

# An Extremal Inequality for Long Markov Chains

Thomas A. Courtade

Department of Electrical Engineering and Computer Sciences  
University of California, Berkeley

Jiantao Jiao

Department of Electrical Engineering  
Stanford University

## Abstract

Let  $\mathbf{X}, \mathbf{Y}$  be jointly Gaussian vectors, and consider random variables  $U, V$  that satisfy the Markov constraint  $U - \mathbf{X} - \mathbf{Y} - V$ . We prove an extremal inequality relating the mutual informations between all  $\binom{4}{2}$  pairs of random variables from the set  $(U, \mathbf{X}, \mathbf{Y}, V)$ . As a first application, we show that the rate region for the two-encoder quadratic Gaussian source coding problem follows as an immediate corollary of the extremal inequality. In a second application, we establish the rate region for a vector-Gaussian source coding problem where Löwner-John ellipsoids are approximated based on rate-constrained descriptions of the data.

## I. INTRODUCTION

In this paper, we prove the following extremal result, which resembles an entropy power inequality:

**Theorem 1.** *For  $n \times n$  positive definite matrices  $\Sigma_X, \Sigma_Z$ , let  $\mathbf{X} \sim N(\mu_X, \Sigma_X)$  and  $\mathbf{Z} \sim N(\mu_Z, \Sigma_Z)$  be independent  $n$ -dimensional Gaussian vectors, and define  $\mathbf{Y} = \mathbf{X} + \mathbf{Z}$ . For any  $U, V$  such that  $U - \mathbf{X} - \mathbf{Y} - V$  form a Markov chain, the following inequality holds:*

$$2^{-\frac{2}{n}(I(\mathbf{Y};U)+I(\mathbf{X};V))} \geq \frac{|\Sigma_X|^{1/n}}{|\Sigma_X + \Sigma_Z|^{1/n}} 2^{-\frac{2}{n}(I(\mathbf{X};U)+I(\mathbf{Y};V))} + 2^{-\frac{2}{n}(I(\mathbf{X};\mathbf{Y})+I(U;V))}. \quad (1)$$

In the simplest case, where  $\mathbf{Y} = \rho\mathbf{X} + \mathbf{Z}$ ,  $\Sigma_X = I_n$  and  $\Sigma_Z = (1 - \rho^2)I_n$ , Theorem 1 implies

$$2^{-\frac{2}{n}(I(\mathbf{Y};U)+I(\mathbf{X};V))} \geq (1 - \rho^2)2^{-\frac{2}{n}I(V;U)} + \rho^2 2^{-\frac{2}{n}(I(\mathbf{X};U)+I(\mathbf{Y};V))}. \quad (2)$$

If  $V$  is degenerate, (2) further simplifies to an inequality shown by Oohama in [1], which proved to be instrumental in establishing the rate-distortion region for the one-helper quadratic Gaussian source coding problem. Together with Oohama's work, the sum-rate constraint established by Wagner *et al.* in their *tour de force* [2] completely characterized the rate-distortion region for the two-encoder quadratic Gaussian source coding problem. It turns out that the sum-rate constraint of Wagner *et al.* can be recovered as an immediate corollary to (2), thus unifying the works of Oohama and Wagner *et al.* under a common inequality. The entire argument is given as follows.

### A. Recovery of the scalar-Gaussian sum-rate constraint

Using the Markov relationship  $U - \mathbf{X} - \mathbf{Y} - V$ , we can rearrange the exponents in (2) to obtain the equivalent inequality

$$2^{-\frac{2}{n}(I(\mathbf{X};U,V)+I(\mathbf{Y};U,V))} \geq 2^{-\frac{2}{n}I(\mathbf{X},\mathbf{Y};U,V)} \left(1 - \rho^2 + \rho^2 2^{-\frac{2}{n}I(\mathbf{X},\mathbf{Y};U,V)}\right). \quad (3)$$

The left- and right-hand sides of (3) are monotone decreasing in  $\frac{1}{n}(I(\mathbf{X};U,V)+I(\mathbf{Y};U,V))$  and  $\frac{1}{n}I(\mathbf{X},\mathbf{Y};U,V)$ , respectively. Therefore, if

$$\frac{1}{n}(I(\mathbf{X};U,V) + I(\mathbf{Y};U,V)) \geq \frac{1}{2} \log \frac{1}{D} \quad \text{and} \quad \frac{1}{n}I(\mathbf{X},\mathbf{Y};U,V) \leq R \quad (4)$$

for some pair  $(R, D)$ , then we have  $D \geq 2^{-2R} (1 - \rho^2 + \rho^2 2^{-2R})$ , which is a quadratic inequality with respect to the term  $2^{-2R}$ . This is easily solved using the quadratic formula to obtain:

$$2^{-2R} \leq \frac{2D}{(1 - \rho^2)\beta(D)} \Rightarrow R \geq \frac{1}{2} \log \frac{(1 - \rho^2)\beta(D)}{2D}, \quad (5)$$

where  $\beta(D) \triangleq 1 + \sqrt{1 + \frac{4\rho^2 D}{(1 - \rho^2)^2}}$ . Note that Jensen's inequality and the maximum-entropy property of Gaussians imply

$$\frac{1}{n}(I(\mathbf{X};U,V) + I(\mathbf{Y};U,V)) \geq \frac{1}{2} \log \frac{1}{\text{mmse}(\mathbf{X}|U,V)\text{mmse}(\mathbf{Y}|U,V)}, \quad (6)$$

where  $\text{mmse}(\mathbf{X}|U,V) \triangleq \frac{1}{n}\|\mathbf{X} - \mathbb{E}[\mathbf{X}|U,V]\|^2$ , and  $\text{mmse}(\mathbf{Y}|U,V)$  is defined similarly. Put  $U = f_x(\mathbf{X})$  and  $V = f_y(\mathbf{Y})$ , where  $f_x : \mathbb{R}^n \rightarrow [1 : 2^{nR_x}]$  and  $f_y : \mathbb{R}^n \rightarrow [1 : 2^{nR_y}]$ . Supposing  $\text{mmse}(\mathbf{X}|U,V) \leq d_x$  and  $\text{mmse}(\mathbf{Y}|U,V) \leq d_y$ , inequalities (4)-(6) together imply

$$R_x + R_y \geq \frac{1}{2} \log \frac{(1 - \rho^2)\beta(d_x d_y)}{2d_x d_y}, \quad (7)$$

which is precisely the sum-rate constraint for the two-encoder quadratic Gaussian source coding problem.

This work was supported by the NSF Center for Science of Information under grant agreement CCF-0939370.

### B. Distributed compression of minimal-volume ellipsoids

Above, recovery of the quadratic Gaussian sum-rate constraint (7) demonstrated the utility of Theorem 1 in proving nontrivial results. Now, we consider a new problem which, to the authors' knowledge, is not a consequence of known results in the open literature. In particular, we study the problem of compressing ellipsoids that cover a set of points which, subject to rate constraints, have approximately minimal volume. Such ellipsoids are similar to Löwner-John ellipsoids, which are defined as the (unique) ellipsoid of minimal volume that covers a finite set of points [3]. These minimum-volume ellipsoids and their approximations play a prominent role in the fields of optimization, data analysis, and computational geometry (e.g., [4]).

To begin, we recall that an  $n$ -dimensional ellipsoid  $\mathcal{E}$  can be parameterized by a positive semidefinite matrix  $A \in \mathbb{R}^{n \times n}$  and a vector  $b \in \mathbb{R}^n$  as follows:

$$\mathcal{E} = \mathcal{E}(A, b) = \{x \in \mathbb{R}^n : \|Ax - b\| \leq 1\}. \quad (8)$$

The volume of  $\mathcal{E}(A, b)$  is related to the determinant of  $A$  by

$$\text{vol}(\mathcal{E}(A, b)) = \frac{c_n}{|A|}, \quad (9)$$

where  $c_n \sim \frac{1}{\sqrt{n\pi}} \left(\frac{2\pi e}{n}\right)^{n/2}$  is the volume of the  $n$ -dimensional unit ball.

Fix  $\rho \in (0, 1)$ , and let  $\{\Sigma_n : n \geq 1\}$  be a sequence of positive definite  $n \times n$  matrices. Suppose  $(\mathbf{X}_1, \mathbf{Y}_1), \dots, (\mathbf{X}_k, \mathbf{Y}_k)$  are  $k$  independent pairs of jointly Gaussian vectors, each equal in distribution to  $(\mathbf{X}, \mathbf{Y})$ , where  $\mathbb{E}[\mathbf{X}\mathbf{X}^T] = \mathbb{E}[\mathbf{Y}\mathbf{Y}^T] = \Sigma_n$ , and  $\mathbb{E}[\mathbf{X}\mathbf{Y}^T] = \rho\Sigma_n$ .

A  $(n, R_x, R_y, \nu_x, \nu_y, k, \Sigma_n, \epsilon)$ -code consists of encoding functions

$$f_x : \mathbb{R}^{kn} \rightarrow \{1, 2, \dots, 2^{knR_x}\} \quad (10)$$

$$f_y : \mathbb{R}^{kn} \rightarrow \{1, 2, \dots, 2^{knR_y}\} \quad (11)$$

and a decoding function

$$\psi : (f_x(\mathbf{X}_1, \dots, \mathbf{X}_k), f_y(\mathbf{Y}_1, \dots, \mathbf{Y}_k)) \mapsto (A_x, A_y, b_x, b_y) \quad (12)$$

such that

$$\max_{1 \leq i \leq k} \Pr\{\mathbf{X}_i \notin \mathcal{E}(A_x, b_x)\} < \epsilon \quad \text{and} \quad \max_{1 \leq i \leq k} \Pr\{\mathbf{Y}_i \notin \mathcal{E}(A_y, b_y)\} < \epsilon, \quad (13)$$

and

$$\left(\text{vol}(\mathcal{E}(A_x, b_x))\right)^{1/n} \leq (1 + \epsilon) c_n^{1/n} \sqrt{n\nu_x |\Sigma_n|^{1/n}} \quad (14)$$

$$\left(\text{vol}(\mathcal{E}(A_y, b_y))\right)^{1/n} \leq (1 + \epsilon) c_n^{1/n} \sqrt{n\nu_y |\Sigma_n|^{1/n}}. \quad (15)$$

We remark that  $\sqrt{n}c_n^{1/n} \rightarrow \sqrt{2\pi e}$  as  $n \rightarrow \infty$  by Stirling's approximation, which explains the normalization factor of  $\sqrt{n}$  in the volume constraint. In particular, (14)-(15) can be replaced with

$$\left(\text{vol}(\mathcal{E}(A_x, b_x))\right)^{1/n} \leq (1 + \epsilon) \sqrt{(2\pi e)\nu_x |\Sigma_n|^{1/n}} \quad (16)$$

$$\left(\text{vol}(\mathcal{E}(A_y, b_y))\right)^{1/n} \leq (1 + \epsilon) \sqrt{(2\pi e)\nu_y |\Sigma_n|^{1/n}}. \quad (17)$$

**Definition 1.** For a sequence  $\{\Sigma_n : n \geq 1\}$  of positive definite  $n \times n$  matrices, a tuple  $(R_x, R_y, \nu_x, \nu_y, k)$  is  $\{\Sigma_n : n \geq 1\}$ -achievable if there exists a sequence of  $(n, R_x, R_y, \nu_x, \nu_y, k, \Sigma_n, \epsilon_n)$  codes satisfying  $\epsilon_n \rightarrow 0$  as  $n \rightarrow \infty$ .

If  $(R_x, R_y, \nu_x, \nu_y, k)$  is a Pareto-optimal  $\{\Sigma_n : n \geq 1\}$ -achievable point, the corresponding ellipsoids  $\mathcal{E}(A_x, b_x), \mathcal{E}(A_y, b_y)$  can be viewed as the best approximations to Löwner-John ellipsoids subject to rate-constrained descriptions of the data. That is, the two ellipsoids cover the  $k$  points observed at their respective encoders, and are (essentially) the minimum-volume such ellipsoids that can be computed from rate-constrained descriptions of the data. The general problem setup is illustrated in Figure 1.

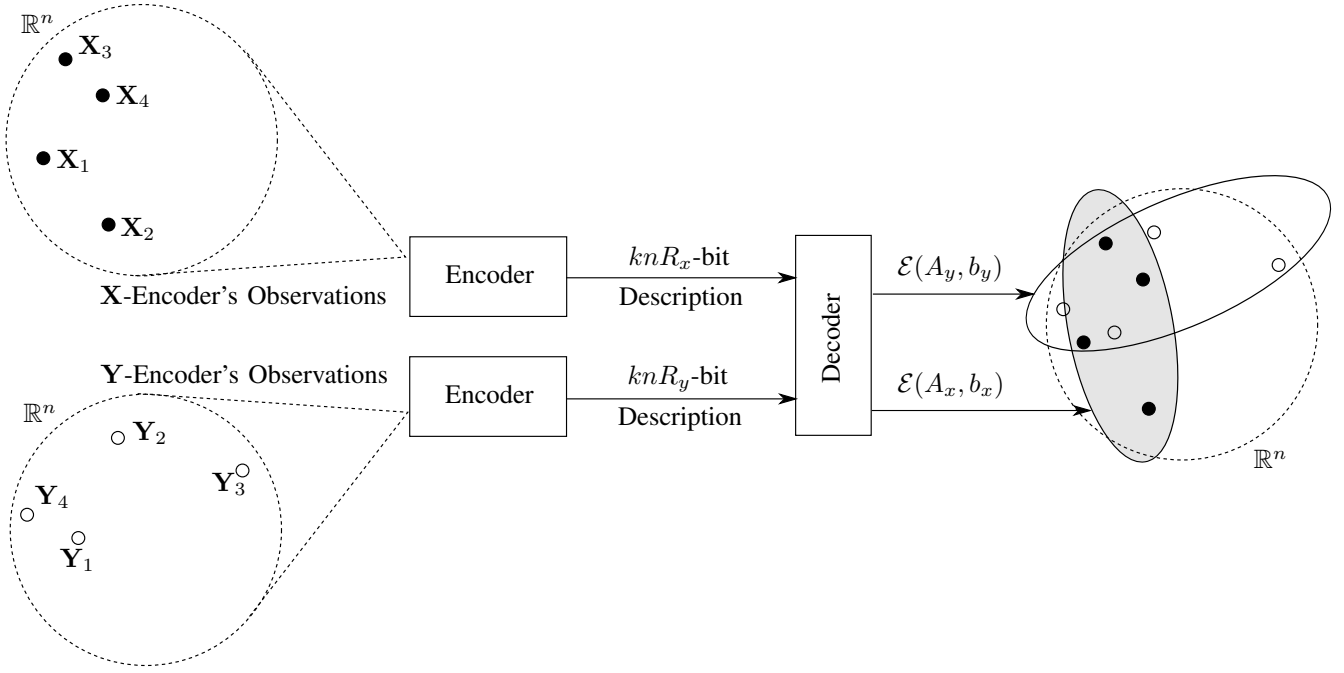


Fig. 1: Computation of covering-ellipsoids from compressed descriptions of the observed data points ( $k = 4$ ). Note the decoder only computes the ellipsoids  $\mathcal{E}(A_x, b_x), \mathcal{E}(A_y, b_y)$ . The data points at the output of the decoder are only shown for reference.

**Theorem 2.** For any sequence  $\{\Sigma_n : n \geq 1\}$  of positive definite  $n \times n$  matrices, a tuple  $(R_x, R_y, \nu_x, \nu_y, k)$  is  $\{\Sigma_n : n \geq 1\}$ -achievable if and only if

$$R_x \geq \frac{1}{2} \log \left[ \frac{1}{\nu_x} (1 - \rho^2 + \rho^2 2^{-2R_y}) \right] \quad (18)$$

$$R_y \geq \frac{1}{2} \log \left[ \frac{1}{\nu_y} (1 - \rho^2 + \rho^2 2^{-2R_x}) \right] \quad (19)$$

$$R_x + R_y \geq \frac{1}{2} \log \frac{(1 - \rho^2) \beta(\nu_x \nu_y)}{2\nu_x \nu_y}, \quad (20)$$

where  $\beta(z) \triangleq 1 + \sqrt{1 + \frac{4\rho^2 z}{(1 - \rho^2)^2}}$ .

**Remark 1.** As we will see, the direct part of Theorem 2 follows from an application of the achievability scheme for the two-encoder quadratic Gaussian source coding problem. However, the converse result does not appear to be a similar consequence since the matrices  $A_x, A_y$  describing the principal axes of the ellipsoids are allowed to depend on the source realizations. Nonetheless, with Theorem 1 at our disposal, the proof of the converse is fairly routine.

Since the primary focus of this paper is on the extremal inequality (1), we defer the proof of Theorem 2 until Appendix A. The remainder of this paper is divided into two parts: a treatment of the scalar version of (1) is given in Section II, and the vector generalization is considered in Section III. Closing remarks are provided in Section IV.

## II. SCALAR SETTING

We begin the journey toward our main result by studying the scalar version of Theorem 1. Most of our effort will carry over to the vector setting, but the notation in the scalar case is less cumbersome. Therefore, for the remainder of this section, we will assume that  $X, Y$  are jointly Gaussian, each with unit variance and correlation  $\rho$ . Our main result in this section is the following rearrangement of (1).

**Theorem 3.** Suppose  $X, Y$  are jointly Gaussian, each with unit variance and correlation  $\rho$ . Then, for any  $U, V$  satisfying  $U - X - Y - V$ , the following inequality holds:

$$2^{-2I(Y;U)} 2^{-2I(X;V|U)} \geq (1 - \rho^2) + \rho^2 2^{-2I(X;U)} 2^{-2I(Y;V|U)}. \quad (21)$$

### A. Proof of Theorem 3

Instead of working directly with inequality (21), it will be convenient to consider a dual form. To this end, for  $\lambda \geq 0$ , define

$$F(\lambda) \triangleq \inf_{U, V: U-X-Y-V} \left\{ I(X; U) - \lambda I(Y; U) + I(Y; V|U) - \lambda I(X; V|U) \right\}. \quad (22)$$

The remainder of this section is devoted to characterizing the function  $F(\lambda)$ . We remark that the infimum in (22) is attained for any  $\lambda$ . The proof of this is routine, and deferred to Appendix B. The bulk of the work ahead is devoted to establishing the existence of valid minimizers  $U, V$  for which  $X|U = u$  is normal for almost every  $u$ .

To accomplish this, we now describe a simple construction that will be used throughout much of the sequel. This construction was first introduced for proving extremal inequalities in [5]. Suppose  $U, X, Y, V$  satisfy the Markov relationship  $U-X-Y-V$ , and consider two independent copies of  $U, X, Y, V$ , which will be denoted by the same variables with subscripts 1 and 2. Define

$$\hat{X}_1 = \frac{X_1 + X_2}{\sqrt{2}} \quad \hat{X}_2 = \frac{X_1 - X_2}{\sqrt{2}}. \quad (23)$$

In a similar manner, define  $\hat{Y}_1, \hat{Y}_2$ . Note that  $(\hat{X}_1, \hat{X}_2, \hat{Y}_1, \hat{Y}_2)$  and  $(X_1, X_2, Y_1, Y_2)$  are equal in distribution. Let  $g: \mathbb{R}^2 \rightarrow \mathbb{R}$  be a one-to-one measurable transformation<sup>1</sup>. Define  $\hat{U} = g(U_1, U_2)$  and  $\hat{V} = g(V_1, V_2)$ .

**Lemma 1.** *If  $U, X, Y, V$  minimize the functional (22), and  $\hat{X}_1, \hat{X}_2, \hat{Y}_1, \hat{Y}_2, \hat{U}, \hat{V}$  are constructed as above, then*

- 1) *For almost every  $y$ ,  $\hat{U}, \hat{X}_1, \hat{Y}_1, \hat{V}$  conditioned on  $\{\hat{Y}_2 = y\}$  is a valid minimizer of (22).*
- 2) *For almost every  $y$ ,  $\hat{U}, \hat{X}_2, \hat{Y}_2, \hat{V}$  conditioned on  $\{\hat{Y}_1 = y\}$  is a valid minimizer of (22).*

*Proof:* Let  $\phi_1$  be such that  $\hat{X}_1, \hat{Y}_1$  are independent of  $\phi_1$  and  $\hat{U} - \hat{X}_1 - \hat{Y}_1 - \hat{V}$  conditioned on  $\phi_1$ . Valid assignments of  $\phi_1$  include any nonempty subset of  $\hat{X}_2, \hat{Y}_2$ . Let  $\phi_2$  be defined similarly.

Now, observe that we can write:

$$2I(X; U) = I(\hat{X}_1, \hat{X}_2; \hat{U}) = I(\hat{X}_1; \hat{U}) + I(\hat{X}_2; \hat{U}|\hat{X}_1) \quad (24)$$

$$= I(\hat{X}_1; \hat{U}) + I(\hat{X}_2; \hat{U}, \hat{X}_1) \quad (25)$$

$$= I(\hat{X}_1; \hat{U}) + I(\hat{X}_2; \hat{U}) + I(\hat{X}_2; \hat{X}_1|\hat{U}) \quad (26)$$

$$= I(\hat{X}_1; \hat{U}|\phi_1) + I(\hat{X}_2; \hat{U}|\phi_2) - I(\hat{X}_1; \phi_1|\hat{U}) - I(\hat{X}_2; \phi_2|\hat{U}) + I(\hat{X}_1; \hat{X}_2|\hat{U}). \quad (27)$$

Similarly,

$$2I(Y; U) = I(\hat{Y}_1; \hat{U}|\phi_1) + I(\hat{Y}_2; \hat{U}|\phi_2) - I(\hat{Y}_1; \phi_1|\hat{U}) - I(\hat{Y}_2; \phi_2|\hat{U}) + I(\hat{Y}_1; \hat{Y}_2|\hat{U}) \quad (28)$$

$$(29)$$

Also, we have

$$2I(Y; V|U) = I(\hat{Y}_1, \hat{Y}_2; \hat{V}|\hat{U}) = I(\hat{Y}_1; \hat{V}|\hat{U}) + I(\hat{Y}_2; \hat{V}|\hat{U}) - I(\hat{Y}_2; \hat{Y}_1|\hat{U}) + I(\hat{Y}_2; \hat{Y}_1|\hat{U}, \hat{V}) \quad (30)$$

$$= I(\hat{Y}_1; \phi_1, \hat{V}|\hat{U}) + I(\hat{Y}_2; \phi_2, \hat{V}|\hat{U}) - I(\hat{Y}_1; \phi_1|\hat{U}, \hat{V}) - I(\hat{Y}_2; \phi_2|\hat{U}, \hat{V}) \quad (31)$$

$$- I(\hat{Y}_2; \hat{Y}_1|\hat{U}) + I(\hat{Y}_2; \hat{Y}_1|\hat{U}, \hat{V})$$

$$= I(\hat{Y}_1; \hat{V}|\hat{U}, \phi_1) + I(\hat{Y}_2; \hat{V}|\hat{U}, \phi_2) \quad (32)$$

$$- I(\hat{Y}_1; \phi_1|\hat{U}, \hat{V}) - I(\hat{Y}_2; \phi_2|\hat{U}, \hat{V}) + I(\hat{Y}_2; \hat{Y}_1|\hat{U}, \hat{V})$$

$$+ I(\hat{Y}_1; \phi_1|\hat{U}) + I(\hat{Y}_2; \phi_2|\hat{U}) - I(\hat{Y}_1; \hat{Y}_2|\hat{U}).$$

And, similarly,

$$2I(X; V|U) = I(\hat{X}_1; \hat{V}|\hat{U}, \phi_1) + I(\hat{X}_2; \hat{V}|\hat{U}, \phi_2) \quad (33)$$

$$- I(\hat{X}_1; \phi_1|\hat{U}, \hat{V}) - I(\hat{X}_2; \phi_2|\hat{U}, \hat{V}) + I(\hat{X}_2; \hat{X}_1|\hat{U}, \hat{V})$$

$$+ I(\hat{X}_1; \phi_1|\hat{U}) + I(\hat{X}_2; \phi_2|\hat{U}) - I(\hat{X}_1; \hat{X}_2|\hat{U}).$$

<sup>1</sup>Every uncountable Polish space is Borel isomorphic to  $\mathbb{R}$  [6].

Assume  $U, V$  minimize the functional (22) subject to the Markov constraint  $U - X - Y - V$ , the existence of such  $U, V$  was established in Lemma 12. Then, combining above, we have

$$2F(\lambda) = I(\hat{X}_1; \hat{U}|\phi_1) + I(\hat{X}_2; \hat{U}|\phi_2) - I(\hat{X}_1; \phi_1|\hat{U}) - I(\hat{X}_2; \phi_2|\hat{U}) + I(\hat{X}_1; \hat{X}_2|\hat{U}) \quad (34)$$

$$\begin{aligned} & - \lambda \left( I(\hat{Y}_1; \hat{U}|\phi_1) + I(\hat{Y}_2; \hat{U}|\phi_2) - I(\hat{Y}_1; \phi_1|\hat{U}) - I(\hat{Y}_2; \phi_2|\hat{U}) + I(\hat{Y}_1; \hat{Y}_2|\hat{U}) \right) \\ & + I(\hat{Y}_1; \hat{V}|\hat{U}, \phi_1) + I(\hat{Y}_2; \hat{V}|\hat{U}, \phi_2) \\ & - I(\hat{Y}_1; \phi_1|\hat{U}, \hat{V}) - I(\hat{Y}_2; \phi_2|\hat{U}, \hat{V}) + I(\hat{Y}_2; \hat{Y}_1|\hat{U}, \hat{V}) \\ & + I(\hat{Y}_1; \phi_1|\hat{U}) + I(\hat{Y}_2; \phi_2|\hat{U}) - I(\hat{Y}_1; \hat{Y}_2|\hat{U}) \\ & - \lambda \left( I(\hat{X}_1; \hat{V}|\hat{U}, \phi_1) + I(\hat{X}_2; \hat{V}|\hat{U}, \phi_2) \right. \\ & \quad \left. - I(\hat{X}_1; \phi_1|\hat{U}, \hat{V}) - I(\hat{X}_2; \phi_2|\hat{U}, \hat{V}) + I(\hat{X}_2; \hat{X}_1|\hat{U}, \hat{V}) \right. \\ & \quad \left. + I(\hat{X}_1; \phi_1|\hat{U}) + I(\hat{X}_2; \phi_2|\hat{U}) - I(\hat{X}_1; \hat{X}_2|\hat{U}) \right) \\ & = I(\hat{X}_1; \hat{U}|\phi_1) - \lambda I(\hat{Y}_1; \hat{U}|\phi_1) + I(\hat{Y}_1; \hat{V}|\hat{U}, \phi_1) - \lambda I(\hat{X}_1; \hat{V}|\hat{U}, \phi_1) \\ & \quad + I(\hat{X}_2; \hat{U}|\phi_2) - \lambda I(\hat{Y}_2; \hat{U}|\phi_2) + I(\hat{Y}_2; \hat{V}|\hat{U}, \phi_2) - \lambda I(\hat{X}_2; \hat{V}|\hat{U}, \phi_2) \\ & \quad - (\lambda + 1)I(\hat{X}_1; \phi_1|\hat{U}) - (\lambda + 1)I(\hat{X}_2; \phi_2|\hat{U}) + (\lambda + 1)I(\hat{X}_1; \hat{X}_2|\hat{U}) \\ & \quad + (\lambda + 1)I(\hat{Y}_1; \phi_1|\hat{U}) + (\lambda + 1)I(\hat{Y}_2; \phi_2|\hat{U}) - (\lambda + 1)I(\hat{Y}_1; \hat{Y}_2|\hat{U}) \\ & \quad - I(\hat{Y}_1; \phi_1|\hat{U}, \hat{V}) - I(\hat{Y}_2; \phi_2|\hat{U}, \hat{V}) + I(\hat{Y}_2; \hat{Y}_1|\hat{U}, \hat{V}) \\ & \quad + \lambda I(\hat{X}_1; \phi_1|\hat{U}, \hat{V}) + \lambda I(\hat{X}_2; \phi_2|\hat{U}, \hat{V}) - \lambda I(\hat{X}_2; \hat{X}_1|\hat{U}, \hat{V}) \end{aligned} \quad (35)$$

$$\begin{aligned} & \geq 2F(\lambda) \\ & \quad - (\lambda + 1)I(\hat{X}_1; \phi_1|\hat{U}) - (\lambda + 1)I(\hat{X}_2; \phi_2|\hat{U}) + (\lambda + 1)I(\hat{X}_1; \hat{X}_2|\hat{U}) \\ & \quad + (\lambda + 1)I(\hat{Y}_1; \phi_1|\hat{U}) + (\lambda + 1)I(\hat{Y}_2; \phi_2|\hat{U}) - (\lambda + 1)I(\hat{Y}_1; \hat{Y}_2|\hat{U}) \\ & \quad - I(\hat{Y}_1; \phi_1|\hat{U}, \hat{V}) - I(\hat{Y}_2; \phi_2|\hat{U}, \hat{V}) + I(\hat{Y}_2; \hat{Y}_1|\hat{U}, \hat{V}) \\ & \quad + \lambda I(\hat{X}_1; \phi_1|\hat{U}, \hat{V}) + \lambda I(\hat{X}_2; \phi_2|\hat{U}, \hat{V}) - \lambda I(\hat{X}_2; \hat{X}_1|\hat{U}, \hat{V}). \end{aligned} \quad (36)$$

The last inequality follows since  $\hat{U} - \hat{X}_1 - \hat{Y}_1 - \hat{V}$  conditioned on  $\phi_1$  is a candidate minimizer of the functional, and same for  $\hat{U} - \hat{X}_2 - \hat{Y}_2 - \hat{V}$  conditioned on  $\phi_2$ . Hence, we can conclude that the following must hold

$$\begin{aligned} & (\lambda + 1)I(\hat{X}_1; \phi_1|\hat{U}) + (\lambda + 1)I(\hat{X}_2; \phi_2|\hat{U}) - (\lambda + 1)I(\hat{X}_1; \hat{X}_2|\hat{U}) \\ & \quad + I(\hat{Y}_1; \phi_1|\hat{U}, \hat{V}) + I(\hat{Y}_2; \phi_2|\hat{U}, \hat{V}) - I(\hat{Y}_2; \hat{Y}_1|\hat{U}, \hat{V}) \\ & \geq (\lambda + 1)I(\hat{Y}_1; \phi_1|\hat{U}) + (\lambda + 1)I(\hat{Y}_2; \phi_2|\hat{U}) - (\lambda + 1)I(\hat{Y}_1; \hat{Y}_2|\hat{U}) \\ & \quad + \lambda I(\hat{X}_1; \phi_1|\hat{U}, \hat{V}) + \lambda I(\hat{X}_2; \phi_2|\hat{U}, \hat{V}) - \lambda I(\hat{X}_2; \hat{X}_1|\hat{U}, \hat{V}). \end{aligned} \quad (37)$$

Now, set  $\phi_1 = \hat{X}_2, \phi_2 = \hat{Y}_1$ . The LHS of (37) is given by

$$\begin{aligned} & (\lambda + 1)I(\hat{X}_1; \phi_1|\hat{U}) + (\lambda + 1)I(\hat{X}_2; \phi_2|\hat{U}) - (\lambda + 1)I(\hat{X}_1; \hat{X}_2|\hat{U}) \\ & \quad + I(\hat{Y}_1; \phi_1|\hat{U}, \hat{V}) + I(\hat{Y}_2; \phi_2|\hat{U}, \hat{V}) - I(\hat{Y}_2; \hat{Y}_1|\hat{U}, \hat{V}) \\ & = (\lambda + 1)I(\hat{X}_1; \hat{X}_2|\hat{U}) + (\lambda + 1)I(\hat{X}_2; \hat{Y}_1|\hat{U}) - (\lambda + 1)I(\hat{X}_1; \hat{X}_2|\hat{U}) \\ & \quad + I(\hat{Y}_1; \hat{X}_2|\hat{U}, \hat{V}) + I(\hat{Y}_2; \hat{Y}_1|\hat{U}, \hat{V}) - I(\hat{Y}_2; \hat{Y}_1|\hat{U}, \hat{V}) \end{aligned} \quad (38)$$

$$= (\lambda + 1)I(\hat{Y}_1; \hat{X}_2|\hat{U}) + I(\hat{Y}_1; \hat{X}_2|\hat{U}, \hat{V}). \quad (39)$$

Also, the RHS of (37) can be expressed as

$$\begin{aligned} & (\lambda + 1)I(\hat{Y}_1; \phi_1|\hat{U}) + (\lambda + 1)I(\hat{Y}_2; \phi_2|\hat{U}) - (\lambda + 1)I(\hat{Y}_1; \hat{Y}_2|\hat{U}) \\ & \quad + \lambda I(\hat{X}_1; \phi_1|\hat{U}, \hat{V}) + \lambda I(\hat{X}_2; \phi_2|\hat{U}, \hat{V}) - \lambda I(\hat{X}_2; \hat{X}_1|\hat{U}, \hat{V}) \\ & = (\lambda + 1)I(\hat{Y}_1; \hat{X}_2|\hat{U}) + (\lambda + 1)I(\hat{Y}_2; \hat{Y}_1|\hat{U}) - (\lambda + 1)I(\hat{Y}_1; \hat{Y}_2|\hat{U}) \\ & \quad + \lambda I(\hat{X}_1; \hat{X}_2|\hat{U}, \hat{V}) + \lambda I(\hat{X}_2; \hat{Y}_1|\hat{U}, \hat{V}) - \lambda I(\hat{X}_2; \hat{X}_1|\hat{U}, \hat{V}) \end{aligned} \quad (40)$$

$$= (\lambda + 1)I(\hat{Y}_1; \hat{X}_2|\hat{U}) + \lambda I(\hat{Y}_1; \hat{X}_2|\hat{U}, \hat{V}). \quad (41)$$

Substituting into (37), we find that  $(\lambda - 1)I(\hat{Y}_1; \hat{X}_2|\hat{U}, \hat{V}) \leq 0 \Rightarrow I(\hat{Y}_1; \hat{X}_2|\hat{U}, \hat{V}) = 0$ . Therefore, (37) is met with equality, and it follows that:

$$F(\lambda) = I(\hat{X}_1; \hat{U}|\hat{X}_2) - \lambda I(\hat{Y}_1; \hat{U}|\hat{X}_2) + I(\hat{Y}_1; \hat{V}|\hat{U}, \hat{X}_2) - \lambda I(\hat{X}_1; \hat{V}|\hat{U}, \hat{X}_2) \quad (42)$$

$$= I(\hat{X}_2; \hat{U}|\hat{Y}_1) - \lambda I(\hat{Y}_2; \hat{U}|\hat{Y}_1) + I(\hat{Y}_2; \hat{V}|\hat{U}, \hat{Y}_1) - \lambda I(\hat{X}_2; \hat{V}|\hat{U}, \hat{Y}_1). \quad (43)$$

Since  $\hat{U}, \hat{X}_1, \hat{Y}_1, \hat{V}$  conditioned on  $\{\hat{Y}_1 = y\}$  is a candidate minimizer of (22), the second assertion of the claim follows. By a symmetric argument, if we set  $\phi_1 = \hat{Y}_2, \phi_2 = \hat{X}_1$ , the roles of the indices are reversed, and we find that

$$F(\lambda) = I(\hat{X}_1; \hat{U}|\hat{Y}_2) - \lambda I(\hat{Y}_1; \hat{U}|\hat{Y}_2) + I(\hat{Y}_1; \hat{V}|\hat{U}, \hat{Y}_2) - \lambda I(\hat{X}_1; \hat{V}|\hat{U}, \hat{Y}_2) \quad (44)$$

$$= I(\hat{X}_2; \hat{U}|\hat{X}_1) - \lambda I(\hat{Y}_2; \hat{U}|\hat{X}_1) + I(\hat{Y}_2; \hat{V}|\hat{U}, \hat{X}_1) - \lambda I(\hat{X}_2; \hat{V}|\hat{U}, \hat{X}_1). \quad (45)$$

This establishes the first assertion of the claim and completes the proof.  $\blacksquare$

**Lemma 2.** *If  $U_\lambda, V_\lambda$  are valid minimizers of the functional (22) for parameter  $\lambda$ , then*

$$I(Y; U_\lambda) + I(X; V_\lambda|U_\lambda) = -F'(\lambda) \quad \text{for a.e. } \lambda. \quad (46)$$

*Proof:* To begin, let  $U_{\lambda+\Delta}, V_{\lambda+\Delta}$  be arbitrary, valid minimizers of the functional (22) for parameter  $\lambda + \Delta$ , and let  $U_{\lambda-\Delta}, V_{\lambda-\Delta}$  be arbitrary, valid minimizers of the functional (22) for parameter  $\lambda - \Delta$ . Next, note that  $F(\lambda)$  is concave and (strictly) monotone decreasing in  $\lambda$ , and hence  $F'(\lambda)$  exists for a.e.  $\lambda$ . Thus, for any  $\Delta > 0$ ,

$$\begin{aligned} \frac{F(\lambda + \Delta) - F(\lambda)}{\Delta} &= \frac{1}{\Delta} \left( I(X; U_{\lambda+\Delta}) + I(Y; V_{\lambda+\Delta}|U_{\lambda+\Delta}) - (\lambda + \Delta) (I(Y; U_{\lambda+\Delta}) + I(X; V_{\lambda+\Delta}|U_{\lambda+\Delta})) \right) \\ &\quad - \frac{1}{\Delta} \left( I(X; U_\lambda) + I(Y; V_\lambda|U_\lambda) - \lambda (I(Y; U_\lambda) + I(X; V_\lambda|U_\lambda)) \right) \\ &= - \left( I(Y; U_{\lambda+\Delta}) + I(X; V_{\lambda+\Delta}|U_{\lambda+\Delta}) \right) \end{aligned} \quad (47)$$

$$\begin{aligned} &\quad + \frac{1}{\Delta} \left( I(X; U_{\lambda+\Delta}) + I(Y; V_{\lambda+\Delta}|U_{\lambda+\Delta}) - \lambda (I(Y; U_{\lambda+\Delta}) + I(X; V_{\lambda+\Delta}|U_{\lambda+\Delta})) \right) \\ &\quad - \frac{1}{\Delta} \left( I(X; U_\lambda) + I(Y; V_\lambda|U_\lambda) - \lambda (I(Y; U_\lambda) + I(X; V_\lambda|U_\lambda)) \right) \\ &\geq - \left( I(Y; U_{\lambda+\Delta}) + I(X; V_{\lambda+\Delta}|U_{\lambda+\Delta}) \right), \end{aligned} \quad (48)$$

where the last inequality follows since  $U_{\lambda+\Delta}, V_{\lambda+\Delta}$  is a candidate minimizer of (22) with parameter  $\lambda$ .

Similarly,

$$\begin{aligned} \frac{F(\lambda) - F(\lambda - \Delta)}{\Delta} &= \frac{1}{\Delta} \left( I(X; U_\lambda) + I(Y; V_\lambda|U_\lambda) - \lambda (I(Y; U_\lambda) + I(X; V_\lambda|U_\lambda)) \right) \\ &\quad - \frac{1}{\Delta} \left( I(X; U_{\lambda-\Delta}) + I(Y; V_{\lambda-\Delta}|U_{\lambda-\Delta}) - (\lambda - \Delta) (I(Y; U_{\lambda-\Delta}) + I(X; V_{\lambda-\Delta}|U_{\lambda-\Delta})) \right) \\ &= - \left( I(Y; U_{\lambda-\Delta}) + I(X; V_{\lambda-\Delta}|U_{\lambda-\Delta}) \right) \end{aligned} \quad (49)$$

$$\begin{aligned} &\quad + \frac{1}{\Delta} \left( I(X; U_\lambda) + I(Y; V_\lambda|U_\lambda) - \lambda (I(Y; U_\lambda) + I(X; V_\lambda|U_\lambda)) \right) \\ &\quad - \frac{1}{\Delta} \left( I(X; U_{\lambda-\Delta}) + I(Y; V_{\lambda-\Delta}|U_{\lambda-\Delta}) - \lambda (I(Y; U_{\lambda-\Delta}) + I(X; V_{\lambda-\Delta}|U_{\lambda-\Delta})) \right) \\ &\leq - \left( I(Y; U_{\lambda-\Delta}) + I(X; V_{\lambda-\Delta}|U_{\lambda-\Delta}) \right), \end{aligned} \quad (50)$$

where the last inequality follows since  $U_{\lambda-\Delta}, V_{\lambda-\Delta}$  is a candidate minimizer of (22) with parameter  $\lambda$ . Recalling concavity of  $F(\lambda)$ , we have shown

$$I(Y; U_{\lambda+\Delta}) + I(X; V_{\lambda+\Delta}|U_{\lambda+\Delta}) \geq -F'(\lambda) \geq I(Y; U_{\lambda-\Delta}) + I(X; V_{\lambda-\Delta}|U_{\lambda-\Delta}). \quad (51)$$

As  $F'$  is monotone and well-defined up to a set of measure zero, we are justified in writing

$$- \lim_{z \rightarrow \lambda^+} F'(z) \geq I(Y; U_\lambda) + I(X; V_\lambda|U_\lambda) \geq - \lim_{z \rightarrow \lambda^-} F'(z). \quad (52)$$

Since  $F'$  is monotone, it is almost everywhere continuous, and so the LHS and RHS above coincide with  $-F'(\lambda)$  for almost every  $\lambda$ .  $\blacksquare$

Since the derivative  $F'(\lambda)$  is just a function of  $F$  itself, and not of a particular minimizer, we have the following

**Corollary 1.** If  $U_\lambda, V_\lambda$  are valid minimizers of the functional (22) for parameter  $\lambda$ , then

$$I(X; U_\lambda|Y) + I(Y; V_\lambda|X) = F(\lambda) - (\lambda - 1)F'(\lambda) \text{ for a.e. } \lambda. \quad (53)$$

*Proof:* Suppose  $U_\lambda, V_\lambda$  are valid minimizers. Then, we can write:

$$F(\lambda) = I(X; U_\lambda) - \lambda I(Y; U_\lambda) + I(Y; V_\lambda|U_\lambda) - \lambda I(X; V_\lambda|U_\lambda) \quad (54)$$

$$= I(X; U_\lambda) - \lambda I(Y; U_\lambda) + I(Y; V_\lambda) - \lambda I(X; V_\lambda) + (\lambda - 1)I(U_\lambda; V_\lambda) \quad (55)$$

$$= I(X; U_\lambda|Y) + I(Y; V_\lambda|X) + (\lambda - 1)(I(U_\lambda; V_\lambda) - I(Y; U_\lambda) - I(X; V_\lambda)) \quad (56)$$

$$= I(X; U_\lambda|Y) + I(Y; V_\lambda|X) + (\lambda - 1)F'(\lambda), \quad (57)$$

where the last line follows from Lemma 2. ■

**Lemma 3.** If  $U, V$  are valid minimizers of the functional (22), and  $\hat{U}, \hat{X}_1, \hat{Y}_1, \hat{X}_2, \hat{Y}_2, \hat{V}$  are constructed as described above, then there exist valid minimizers  $\tilde{U}, \tilde{V}$  such that

$$I(X; U|Y) \geq I(X; \tilde{U}|Y) + \frac{1}{2}I(\hat{X}_1; \hat{X}_2|\hat{U}, \hat{Y}_1, \hat{Y}_2). \quad (58)$$

*Proof:* To begin, note that:

$$I(\hat{X}_1; \hat{X}_2|\hat{U}, \hat{Y}_1, \hat{Y}_2) = I(\hat{X}_1; \hat{X}_2, \hat{U}|\hat{Y}_1, \hat{Y}_2) - I(\hat{X}_1; \hat{U}|\hat{Y}_1, \hat{Y}_2) \quad (59)$$

$$= I(\hat{X}_1; \hat{U}|\hat{Y}_1, \hat{Y}_2, \hat{X}_2) - I(\hat{X}_1; \hat{U}|\hat{Y}_1, \hat{Y}_2) \quad (60)$$

$$= I(\hat{X}_1, \hat{X}_2; \hat{U}|\hat{Y}_1, \hat{Y}_2) - I(\hat{X}_1; \hat{U}|\hat{Y}_1, \hat{Y}_2) - I(\hat{X}_2; \hat{U}|\hat{Y}_1, \hat{Y}_2) \quad (61)$$

$$= 2I(X; U|Y) - I(\hat{X}_1; \hat{U}|\hat{Y}_1, \hat{Y}_2) - I(\hat{X}_2; \hat{U}|\hat{Y}_1, \hat{Y}_2). \quad (62)$$

Thus, without loss of generality (relabeling indices 1 and 2 if necessary), we can assume

$$I(X; U|Y) \geq I(\hat{X}_1; \hat{U}|\hat{Y}_1, \hat{Y}_2) + \frac{1}{2}I(\hat{X}_1; \hat{X}_2|\hat{U}, \hat{Y}_1, \hat{Y}_2). \quad (63)$$

Lemma 1 asserts that, for almost every  $y$ , the tuple  $\hat{U}, \hat{X}_1, \hat{Y}_1, \hat{V}$  conditioned on  $\{\hat{Y}_2 = y\}$  is a valid minimizer of (22). Hence, there must exist a  $y^*$  such that

$$I(X; U|Y) \geq I(\hat{X}_1; \hat{U}|\hat{Y}_1, \hat{Y}_2 = y^*) + \frac{1}{2}I(\hat{X}_1; \hat{X}_2|\hat{U}, \hat{Y}_1, \hat{Y}_2), \quad (64)$$

and  $\hat{U}, \hat{X}_1, \hat{Y}_1, \hat{V}$  conditioned on  $\{\hat{Y}_2 = y^*\}$  is a valid minimizer of (22). Therefore, the claim follows by letting  $\tilde{U}, X, Y, \tilde{V}$  be equal in distribution to  $\hat{U}, \hat{X}_1, \hat{Y}_1, \hat{V}$  conditioned on  $\{\hat{Y}_2 = y^*\}$ . ■

**Corollary 2.** There exist  $U, V$  which are valid minimizers of the functional (22), and satisfy

$$I(\hat{X}_1; \hat{X}_2|\hat{U}, \hat{Y}_1, \hat{Y}_2) = 0, \quad (65)$$

where  $\hat{U}, \hat{X}_1, \hat{Y}_1, \hat{X}_2, \hat{Y}_2, \hat{V}$  are constructed as described above.

*Proof:* Applying Lemma 3, we can inductively construct a sequence of valid minimizers  $\{U^{(k)}, X, Y, V^{(k)}\}_{k \geq 1}$  which satisfy

$$I(X; U^{(k)}|Y) \geq I(X; U^{(k+1)}|Y) + \frac{1}{2}I(\hat{X}_1; \hat{X}_2|\hat{U}^{(k)}, \hat{Y}_1, \hat{Y}_2) \text{ for } k = 1, 2, \dots, \quad (66)$$

where  $\hat{U}^{(k)}, \hat{X}_1, \hat{Y}_1, \hat{X}_2, \hat{Y}_2, \hat{V}^{(k)}$  are constructed from two independent copies of  $U^{(k)}, X, Y, V^{(k)}$ . By Corollary 1, we must also have

$$I(X; U^{(k)}|Y) + I(Y; V^{(k)}|X) = F(\lambda) - (\lambda - 1)F'(\lambda) \quad (67)$$

for all  $k = 1, 2, \dots$ . Therefore, for any  $n$ , we have:

$$F(\lambda) - (\lambda - 1)F'(\lambda) = \frac{1}{n} \sum_{k=1}^n I(X; U^{(k)}|Y) + I(Y; V^{(k)}|X) \quad (68)$$

$$\geq \frac{1}{n} \sum_{k=2}^n \left( I(X; U^{(k)}|Y) + I(Y; V^{(k)}|X) \right) + \frac{1}{n} \left( I(X; U^{(n+1)}|Y) + I(Y; V^{(1)}|X) \right) \quad (69)$$

$$+ \frac{1}{2n} \sum_{k=1}^n I(\hat{X}_1; \hat{X}_2|\hat{U}^{(k)}, \hat{Y}_1, \hat{Y}_2) \geq \frac{n-1}{n} \left( F(\lambda) - (\lambda - 1)F'(\lambda) \right) + \frac{1}{2n} \sum_{k=1}^n I(\hat{X}_1; \hat{X}_2|\hat{U}^{(k)}, \hat{Y}_1, \hat{Y}_2), \quad (70)$$

and thus

$$\sum_{k=1}^n I(\hat{X}_1; \hat{X}_2|\hat{U}^{(k)}, \hat{Y}_1, \hat{Y}_2) \leq 2 \left( F(\lambda) - (\lambda - 1)F'(\lambda) \right). \quad (71)$$

Hence, the sum on the LHS of (71) must converge as  $n \rightarrow \infty$ , implying

$$\lim_{k \rightarrow \infty} I(\hat{X}_1; \hat{X}_2|\hat{U}^{(k)}, \hat{Y}_1, \hat{Y}_2) = 0. \quad (72)$$

Arguing as in the proof of Lemma 12 in Appendix B, we can conclude that there exists an optimizer  $U, V$  for which  $I(\hat{X}_1; \hat{X}_2|\hat{U}, \hat{Y}_1, \hat{Y}_2)$  is exactly zero. ■

**Lemma 4.** [7] Let  $\mathbf{A}_1$  and  $\mathbf{A}_2$  be mutually independent  $n$ -dimensional random vectors. If  $\mathbf{A}_1 + \mathbf{A}_2$  is independent of  $\mathbf{A}_1 - \mathbf{A}_2$ , then  $\mathbf{A}_1$  and  $\mathbf{A}_2$  are normally distributed.

**Corollary 3.** There exist optimizers  $U, V$  such that  $X|\{U = u\}$  is Gaussian for a.e.  $u$ .

*Proof:* By construction and Corollary 2, we can conclude that there exist optimizers  $U, V$  for which

$$I(X_1; X_2|U_1, U_2, Y_1, Y_2) = I(\hat{X}_1; \hat{X}_2|U_1, U_2, Y_1, Y_2) = 0. \quad (73)$$

Therefore, by Lemma 4, there exist optimizers  $U, V$  such that  $X|\{U, Y = u, y\}$  is Gaussian for a.e.  $u, y$ .

Letting  $P(x, y, u, v)$  denote the joint distribution of the above  $X, Y, U, V$ , we can use Markovity to write:

$$P(x, y, u, v) = P(u)P(y|u)P(x|u, y)P(v|y) \quad (74)$$

$$= P(u)P(x|u)P(y|x)P(v|y). \quad (75)$$

Taking logarithms and rearranging, we have the identity

$$\log(P(x|u)) = \log(P(y|u)) + \log(P(x|u, y)) - \log(P(y|x)). \quad (76)$$

Since  $X|\{U, Y = u, y\}$  is Gaussian for a.e.  $u, y$ , and  $X, Y$  are jointly Gaussian by assumption, the RHS of (76) is a quadratic function of  $x$  for a.e.  $u, y$ . Hence,  $\log(P(x|u))$  is quadratic in  $x$  for a.e.  $u$ , and the claim follows. ■

**Lemma 5.** [1] For any  $U$  satisfying  $U - X - Y$ , the following inequality holds:

$$2^{-2I(Y;U)} \geq 1 - \rho^2 + \rho^2 2^{-2I(X;U)}. \quad (77)$$

*Proof:* Consider any  $U$  satisfying  $U - X - Y$ . Let  $Y_u, X_u$  denote the random variables  $X, Y$  conditioned on  $U = u$ . By Markovity and definition of  $X, Y$ , we have that  $Y_u = \rho X_u + Z$ , where  $Z \sim N(0, 1 - \rho^2)$  is independent of  $X_u$ . Hence, the conditional entropy power inequality implies that

$$2^{2h(Y|U)} \geq \rho^2 2^{2h(X|U)} + 2\pi e(1 - \rho^2) = 2\pi e \rho^2 2^{-2I(X;U)} + 2\pi e(1 - \rho^2).$$

From here, the lemma easily follows. ■

**Lemma 6.**

$$\inf_{U: U-X-Y} \left\{ I(X;U) - \lambda I(Y;U) \right\} = \begin{cases} \frac{1}{2} \left[ \log \left( \frac{\rho^2(\lambda-1)}{1-\rho^2} \right) - \lambda \log \left( \frac{\lambda-1}{\lambda(1-\rho^2)} \right) \right] & \text{If } \lambda \geq 1/\rho^2 \\ 0 & \text{If } 0 \leq \lambda \leq 1/\rho^2. \end{cases} \quad (78)$$

*Proof:* The claim follows from Lemma 5 and Lemma 14 in Appendix C by identifying  $a_1 \leftarrow \rho^2$  and  $a_2 \leftarrow (1 - \rho^2)$ . ■



**Lemma 7.**

$$F(\lambda) = \inf_{U: U-X-Y} \left\{ I(X; U) - \lambda I(Y; U) \right\}. \quad (79)$$

*Proof:* We will assume  $\lambda \geq 1/\rho^2$ . The claim that  $F(\lambda) = 0$  in the complementary case will then follow immediately by monotonicity of  $F(\lambda)$ . To this end, let  $U, V$  be optimizers such that  $X|U = u$  is Gaussian for a.e.  $u$ . The existence of such  $U, V$  is guaranteed by Corollary 3. Let  $X_u, Y_u$  denote the random variables  $X, Y$  conditioned on  $U = u$ . By Markovity,  $X_u, Y_u$  are jointly Gaussian with

$$Y_u = \rho X_u + Z, \quad (80)$$

where  $Z \sim N(0, 1 - \rho^2)$  is independent of  $X_u$ . Letting  $\sigma_u^2$  be the variance of  $X_u$ , the variance of  $Y_u$  is  $\rho^2 \sigma_u^2 + (1 - \rho^2)$ . Moreover, the squared linear correlation of  $X_u$  and  $Y_u$  is given by

$$\rho_u^2 \triangleq \frac{\rho^2 \sigma_u^2}{\rho^2 \sigma_u^2 + (1 - \rho^2)}. \quad (81)$$

By Lemma 6,

$$\inf_{V: V-Y_u-X_u} \left\{ I(Y_u; V) - \lambda I(X_u; V) \right\} = \frac{1}{2} \left[ \log \left( \frac{\rho_u^2 (\lambda - 1)}{1 - \rho_u^2} \right) - \lambda \log \left( \frac{\lambda - 1}{\lambda(1 - \rho_u^2)} \right) \right] \quad (82)$$

whenever  $\lambda \geq 1/\rho_u^2$ , and the infimum is equal to zero otherwise.

By definition, we have

$$F(\lambda) = I(X; U) - \lambda I(Y; U) + I(Y; V|U) - \lambda I(X; V|U) \quad (83)$$

$$= \int \left( h(X) - h(X|u) - \lambda(h(Y) - h(Y|U = u)) + I(Y; V|U = u) - \lambda I(X; V|U = u) \right) dP_U(u) \quad (84)$$

$$= \int \left( -\frac{1}{2} \log \sigma_u^2 + \frac{\lambda}{2} \log(\rho^2 \sigma_u^2 + (1 - \rho^2)) + I(Y; V|U = u) - \lambda I(X; V|U = u) \right) dP_U(u). \quad (85)$$

If  $\lambda \geq 1/\rho_u^2$ , we can apply (82) to bound the integrand in (85) as follows

$$\begin{aligned} & -\frac{1}{2} \log \sigma_u^2 + \frac{\lambda}{2} \log(\rho^2 \sigma_u^2 + (1 - \rho^2)) + I(Y; V|U = u) - \lambda I(X; V|U = u) \\ & \geq -\frac{1}{2} \log \sigma_u^2 + \frac{\lambda}{2} \log(\rho^2 \sigma_u^2 + (1 - \rho^2)) + \frac{1}{2} \left[ \log \left( \frac{\rho_u^2 (\lambda - 1)}{1 - \rho_u^2} \right) - \lambda \log \left( \frac{\lambda - 1}{\lambda(1 - \rho_u^2)} \right) \right] \end{aligned} \quad (86)$$

$$= \frac{1}{2} \left[ \log \left( \frac{\rho^2 (\lambda - 1)}{1 - \rho^2} \right) - \lambda \log \left( \frac{\lambda - 1}{\lambda(1 - \rho^2)} \right) \right]. \quad (87)$$

On the other hand, if  $\lambda \leq 1/\rho_u^2$ , then we can bound the integrand in (85) by

$$\begin{aligned} & -\frac{1}{2} \log \sigma_u^2 + \frac{\lambda}{2} \log(\rho^2 \sigma_u^2 + (1 - \rho^2)) + I(Y; V|U = u) - \lambda I(X; V|U = u) \\ & \geq -\frac{1}{2} \log \sigma_u^2 + \frac{\lambda}{2} \log(\rho^2 \sigma_u^2 + (1 - \rho^2)) \end{aligned} \quad (88)$$

$$\geq \frac{1}{2} \left[ \log \left( \frac{\rho^2 (\lambda - 1)}{1 - \rho^2} \right) - \lambda \log \left( \frac{\lambda - 1}{\lambda(1 - \rho^2)} \right) \right], \quad (89)$$

where the final inequality follows since  $\lambda \leq 1/\rho_u^2 \Rightarrow \sigma_u^2 \leq \frac{1-\rho^2}{\rho^2(\lambda-1)}$ , and  $-\frac{1}{2} \log \sigma_u^2 + \frac{\lambda}{2} \log(\rho^2 \sigma_u^2 + (1 - \rho^2))$  is monotone decreasing in  $\sigma_u^2$  for  $\sigma_u^2 \leq \frac{1-\rho^2}{\rho^2(\lambda-1)}$ . Therefore, we have established the inequality

$$F(\lambda) \geq \frac{1}{2} \left[ \log \left( \frac{\rho^2 (\lambda - 1)}{1 - \rho^2} \right) - \lambda \log \left( \frac{\lambda - 1}{\lambda(1 - \rho^2)} \right) \right]. \quad (90)$$

The definition of  $F(\lambda)$  together with Lemma 6 implies the reverse inequality, completing the proof.  $\blacksquare$

Since (78) is a dual characterization of the inequality (77), we have proved Theorem 3.

**Remark 2.** Although Lemma 7 implies that the functional (22) is minimized when either  $U$  or  $V$  is degenerate, there are also minimizers for which this is not the case. For example, if  $-1 \leq \rho_u, \rho_v \leq 1$  satisfy

$$(1 - \rho^2)(1 - \rho^2 \rho_u^2 \rho_v^2) = \rho^2 (\lambda - 1)(1 - \rho_u^2)(1 - \rho_v^2), \quad (91)$$

then  $U, V$  defined according to

$$U = \rho_u X + Z_u \quad (92)$$

$$V = \rho_v Y + Z_v, \quad (93)$$

where  $Z_u \sim N(0, 1 - \rho_u^2)$  and  $Z_v \sim N(0, 1 - \rho_v^2)$  are independent of everything else, also minimize (22).

### III. VECTOR SETTING

Now, we turn our attention to the vector case. Throughout the remainder of this section, let  $\Sigma_X, \Sigma_Z$  be positive definite  $n \times n$  matrices. Suppose  $\mathbf{X} \sim N(\mu_X, \Sigma_X)$  and  $\mathbf{Z} \sim N(\mu_Z, \Sigma_Z)$  are independent  $n$ -dimensional Gaussian vectors, and define  $\mathbf{Y} = \mathbf{X} + \mathbf{Z}$ . We recall the statement of Theorem 1 here, along with conditions for equality:

**Theorem 1.** For any  $U, V$  such that  $U - \mathbf{X} - \mathbf{Y} - V$ ,

$$2^{-\frac{2}{n}(I(\mathbf{Y};U)+I(\mathbf{X};V|U))} \geq \frac{|\Sigma_X|^{1/n}}{|\Sigma_X + \Sigma_Z|^{1/n}} 2^{-\frac{2}{n}(I(\mathbf{X};U)+I(\mathbf{Y};V|U))} + 2^{-\frac{2}{n}I(\mathbf{X};\mathbf{Y})}. \quad (94)$$

Moreover, equality holds iff  $\mathbf{X}|\{U = u\} \sim N(\mu_u, \Sigma_{X|U})$  for all  $u$ , where  $\mu_u \triangleq \mathbb{E}[\mathbf{X}|U = u]$ , and  $\Sigma_{X|U}$  is proportional to  $\Sigma_Z$ .

#### A. Proof of Theorem 1

Instead of working directly with inequality (94), it will be convenient to consider a dual form. As before, for  $\lambda \geq 0$ , define

$$\mathbf{F}(\lambda) \triangleq \inf_{U, V: U - \mathbf{X} - \mathbf{Y} - V} \left\{ I(\mathbf{X}; U) - \lambda I(\mathbf{Y}; U) + I(\mathbf{Y}; V|U) - \lambda I(\mathbf{X}; V|U) \right\} \quad (95)$$

The remainder of this section is devoted to bounding the function  $\mathbf{F}(\lambda)$ .

To begin, we remark that the extension of the results up to Lemma 5 for the scalar setting immediately generalize to the present vector case by repeating the proofs verbatim. Namely, we have the key observation:

**Corollary 4.** There exist  $U, V$  which minimize the functional (95) such that  $\mathbf{X}|\{U = u\}$  is Gaussian for a.e.  $u$ .

Therefore, we pick up at this point and prove a vector version of Lemma 5.

**Lemma 8.** For any  $U$  such that  $U - \mathbf{X} - \mathbf{Y}$ ,

$$2^{-2I(\mathbf{Y};U)/n} \geq \frac{|\Sigma_X|^{1/n}}{|\Sigma_X + \Sigma_Z|^{1/n}} 2^{-2I(\mathbf{X};U)/n} + 2^{-2I(\mathbf{X};\mathbf{Y})/n}. \quad (96)$$

Moreover, equality holds iff  $\mathbf{X}|\{U = u\} \sim N(\mu_u, \Sigma)$  for all  $u$ , where  $\mu_u \triangleq \mathbb{E}[\mathbf{X}|U = u]$ , and  $\Sigma \propto \Sigma_Z$ .

*Proof:* Without loss of generality, we can assume  $\mu_X = \mu_Z = 0$ . Note that  $\Sigma_Y = \Sigma_X + \Sigma_Z$ . We can diagonalize the covariance matrices as

$$\Sigma_X = U_X \Lambda_X U_X^T \quad (97)$$

$$\Sigma_Y = U_Y \Lambda_Y U_Y^T, \quad (98)$$

where  $U_X$  and  $U_Y$  are unitary. Define  $\mathbf{Y}' = \Lambda_Y^{-1/2} U_Y^T \mathbf{Y}$ , implying  $\mathbf{Y}' \sim N(0, I_n)$  and

$$\mathbf{Y}' = \Lambda_Y^{-1/2} U_Y^T \mathbf{X} + \Lambda_Y^{-1/2} U_Y^T \mathbf{Z}. \quad (99)$$

Further, define  $\mathbf{X}' = \Lambda_X^{-1/2} U_X^T \mathbf{X}$ , so that  $\mathbf{X}' \sim N(0, I_n)$  and

$$\mathbf{Y}' = \Lambda_Y^{-1/2} U_Y^T U_X \Lambda_X^{1/2} \mathbf{X}' + \Lambda_Y^{-1/2} U_Y^T \mathbf{Z} = B \mathbf{X}' + W, \quad (100)$$

where  $B \triangleq \Lambda_Y^{-1/2} U_Y^T U_X \Lambda_X^{1/2}$  and  $W \triangleq \Lambda_Y^{-1/2} U_Y^T \mathbf{Z}$ , implying  $W \sim N(0, \Lambda_Y^{-1/2} U_Y^T \Sigma_Z U_Y \Lambda_Y^{-1/2})$ . For any  $U$  such that  $U - \mathbf{X} - \mathbf{Y}$ ,  $\mathbf{X}'$  and  $W$  are independent given  $U$ . Thus, the conditional entropy power inequality implies

$$2^{2h(\mathbf{Y}'|U)/n} \geq 2^{2h(B\mathbf{X}'|U)/n} + 2^{2h(W|U)/n} \quad (101)$$

$$= |B|^{2/n} 2^{2h(\mathbf{X}'|U)/n} + 2\pi e |\Lambda_Y^{-1/2} U_Y^T \Sigma_Z U_Y \Lambda_Y^{-1/2}|^{1/n} \quad (102)$$

$$= |\Lambda_Y^{-1/2} U_Y^T U_X \Lambda_X^{1/2}|^{2/n} 2^{2h(\mathbf{X}'|U)/n} + 2\pi e \frac{|\Sigma_Z|^{1/n}}{|\Sigma_Y|^{1/n}} \quad (103)$$

$$= \frac{|\Sigma_X|^{1/n}}{|\Sigma_Y|^{1/n}} 2^{2h(\mathbf{X}'|U)/n} + 2\pi e \frac{|\Sigma_Z|^{1/n}}{|\Sigma_Y|^{1/n}}. \quad (104)$$

Multiplying both sides by  $2^{-2h(\mathbf{Y}')/n} = 2^{-2h(\mathbf{X}')/n} = \frac{1}{2\pi e}$ , we obtain:

$$2^{-2I(\mathbf{Y}';U)/n} \geq \frac{|\Sigma_X|^{1/n}}{|\Sigma_Y|^{1/n}} 2^{-2I(\mathbf{X}';U)/n} + \frac{|\Sigma_Z|^{1/n}}{|\Sigma_Y|^{1/n}}. \quad (105)$$

Since mutual information is invariant under one-to-one transformations of support, we also have

$$2^{-2I(\mathbf{Y};U)/n} \geq \frac{|\Sigma_X|^{1/n}}{|\Sigma_Y|^{1/n}} 2^{-2I(\mathbf{X};U)/n} + \frac{|\Sigma_Z|^{1/n}}{|\Sigma_Y|^{1/n}} \quad (106)$$

$$= \frac{|\Sigma_X|^{1/n}}{|\Sigma_X + \Sigma_Z|^{1/n}} 2^{-2I(\mathbf{X};U)/n} + 2^{-2I(\mathbf{X};\mathbf{Y})/n}. \quad (107)$$

The condition for equality follows from the necessary conditions for equality in the conditional entropy power inequality. ■

**Lemma 9.** Let  $U$  be such that  $U - \mathbf{X} - \mathbf{Y}$ .

1) If  $\lambda \geq 1 + |\Sigma_X^{-1}\Sigma_Z|^{1/n}$ , then

$$I(\mathbf{X};U) - \lambda I(\mathbf{Y};U) \geq \frac{n}{2} \left[ \log \left( \frac{|\Sigma_X|^{1/n}(\lambda - 1)}{|\Sigma_Z|^{1/n}} \right) - \lambda \log \left( \frac{|\Sigma_X + \Sigma_Z|^{1/n}(\lambda - 1)}{|\Sigma_Z|^{1/n}\lambda} \right) \right]. \quad (108)$$

2) If  $0 \leq \lambda \leq 1 + |\Sigma_X^{-1}\Sigma_Z|^{1/n}$ , then

$$I(\mathbf{X};U) - \lambda I(\mathbf{Y};U) \geq -\frac{\lambda n}{2} \log \left( \frac{|\Sigma_X + \Sigma_Z|^{1/n}}{|\Sigma_X|^{1/n} + |\Sigma_Z|^{1/n}} \right). \quad (109)$$

*Proof:* The claim follows from Lemma 14 in Appendix C by identifying  $a_1 \leftarrow \frac{|\Sigma_X|^{1/n}}{|\Sigma_X + \Sigma_Z|^{1/n}}$  and  $a_2 \leftarrow 2^{-2I(\mathbf{X};\mathbf{Y})/n} = \frac{|\Sigma_Z|^{1/n}}{|\Sigma_X + \Sigma_Z|^{1/n}}$ . The hypothesis that  $a_1 + a_2 \leq 1$  is satisfied since the Minkowski determinant theorem [8] asserts that  $|\Sigma_X|^{1/n} + |\Sigma_Z|^{1/n} \leq |\Sigma_X + \Sigma_Z|^{1/n}$ . ■

Note that the lower bound (108) is achieved if  $U$  can be chosen such that  $\mathbf{X}|\{U = u\} \sim N(\mu_u, \Sigma_{X|U})$  for each  $u$ , and  $\Sigma_{X|U} = \alpha \Sigma_Z$ . In this case

$$\frac{1}{n} I(\mathbf{X};U) - \frac{\lambda}{n} I(\mathbf{Y};U) = \frac{1}{2n} \log \left( \frac{|\Sigma_X|}{|\Sigma_{X|U}|} \right) - \frac{\lambda}{2n} \log \left( \frac{|\Sigma_X + \Sigma_Z|}{|\Sigma_{X|U} + \Sigma_Z|} \right) \quad (110)$$

$$= \frac{1}{2n} \log \left( \frac{|\Sigma_X|}{\alpha^n |\Sigma_Z|} \right) - \frac{\lambda}{2n} \log \left( \frac{|\Sigma_X + \Sigma_Z|}{(1 + \alpha)^n |\Sigma_Z|} \right) \quad (111)$$

$$= \frac{1}{2} \left[ \log \left( \frac{|\Sigma_X|^{1/n}(\lambda - 1)}{|\Sigma_Z|^{1/n}} \right) - \lambda \log \left( \frac{|\Sigma_X + \Sigma_Z|^{1/n}(\lambda - 1)}{|\Sigma_Z|^{1/n}\lambda} \right) \right], \quad (112)$$

where we set  $\alpha = \frac{1}{\lambda - 1}$  to arrive at the final equality. The lower bound (109) is only attainable if  $\Sigma_X$  and  $\Sigma_Z$  are proportional. In this case, the RHS of (109) is precisely zero.

**Lemma 10.**

1) If  $\lambda \geq 1 + |\Sigma_X^{-1}\Sigma_Z|^{1/n}$ , then

$$\mathbf{F}(\lambda) \geq \frac{n}{2} \left[ \log \left( \frac{|\Sigma_X|^{1/n}(\lambda - 1)}{|\Sigma_Z|^{1/n}} \right) - \lambda \log \left( \frac{|\Sigma_X + \Sigma_Z|^{1/n}(\lambda - 1)}{|\Sigma_Z|^{1/n}\lambda} \right) \right]. \quad (113)$$

2) If  $0 \leq \lambda \leq 1 + |\Sigma_X^{-1}\Sigma_Z|^{1/n}$ , then

$$\mathbf{F}(\lambda) \geq -\frac{\lambda n}{2} \log \left( \frac{|\Sigma_X + \Sigma_Z|^{1/n}}{|\Sigma_X|^{1/n} + |\Sigma_Z|^{1/n}} \right). \quad (114)$$

*Proof:* Similar to the scalar setting, it is sufficient to restrict our attention to the setting where  $\lambda \geq 1 + |\Sigma_X^{-1}\Sigma_Z|^{1/n}$ . Therefore, we assume this throughout the proof.

Let  $U, V$  be valid minimizers of (95), where  $\mathbf{X}|\{U = u\}$  is Gaussian for a.e.  $u$ . The existence of such  $U, V$  is guaranteed by Corollary 4. By definition, we have

$$\mathbf{F}(\lambda) = I(\mathbf{X};U) - \lambda I(\mathbf{Y};U) + I(\mathbf{Y};V|U) - \lambda I(\mathbf{X};V|U) \quad (115)$$

$$= \int \left( h(\mathbf{X}) - h(\mathbf{X}|u) - \lambda(h(\mathbf{Y}) - h(\mathbf{Y}|U = u)) + I(\mathbf{Y};V|U = u) - \lambda I(\mathbf{X};V|U = u) \right) dP_U(u) \quad (116)$$

Let  $\mathbf{X}_u$  and  $\mathbf{Y}_u$  denote the random variables  $\mathbf{X}, \mathbf{Y}$  conditioned on  $\{U = u\}$ . Suppose  $\mathbf{X}_u \sim N(\mu_u, \Sigma_{X_u})$ . Then  $\mathbf{X}_u, \mathbf{Y}_u$  are jointly normal, and we can write  $\mathbf{X}_u = B\mathbf{Y}_u + W$ , where  $W \sim N(0, \Sigma_{X_u} - \Sigma_{X_u Y_u} \Sigma_{Y_u}^{-1} \Sigma_{Y_u X_u})$  is independent of  $\mathbf{Y}_u$ , and  $B = \Sigma_{X_u Y_u} \Sigma_{Y_u}^{-1}$ . Note that Markovity implies  $\mathbf{Y}_u = \mathbf{X}_u + \mathbf{Z}$ , so that  $\Sigma_{X_u Y_u} = \Sigma_{Y_u X_u} = \Sigma_{X_u}$ , which simplifies  $\Sigma_{X_u Y_u} \Sigma_{Y_u}^{-1} \Sigma_{Y_u X_u}$  to  $\Sigma_{X_u} \Sigma_{Y_u}^{-1} \Sigma_{X_u}$ , and also implies  $\Sigma_{Y_u} - \Sigma_{X_u} = \Sigma_Z$ .

Suppose  $\Sigma_{X_u}$  is such that  $\lambda \geq 1 + \frac{|\Sigma_{X_u} - \Sigma_{X_u} \Sigma_{Y_u}^{-1} \Sigma_{X_u}|^{1/n}}{|\Sigma_{X_u} \Sigma_{Y_u}^{-1} \Sigma_{X_u}|^{1/n}} = 1 + \frac{|\Sigma_Z|^{1/n}}{|\Sigma_{X_u}|^{1/n}}$ . Then, Lemma 9 allows us to lower bound the integrand in (116) as:

$$h(\mathbf{X}) - h(\mathbf{X}|u) - \lambda(h(\mathbf{Y}) - h(\mathbf{Y}|u)) + I(\mathbf{Y}; V|U = u) - \lambda I(\mathbf{X}; V|U = u) \geq \frac{n}{2} \left[ \log \left( \frac{|\Sigma_X|^{1/n}}{|\Sigma_{X_u}|^{1/n}} \right) - \lambda \log \left( \frac{|\Sigma_Y|^{1/n}}{|\Sigma_{Y_u}|^{1/n}} \right) \right] \quad (117)$$

$$+ \frac{n}{2} \left[ \log \left( \frac{|B \Sigma_{Y_u} B^T|^{1/n} (\lambda - 1)}{|\Sigma_W|^{1/n}} \right) - \lambda \log \left( \frac{|B \Sigma_{Y_u} B^T + \Sigma_W|^{1/n} (\lambda - 1)}{|\Sigma_W|^{1/n} \lambda} \right) \right] = \frac{n}{2} \left[ \log \left( \frac{|\Sigma_X|^{1/n}}{|\Sigma_{X_u}|^{1/n}} \right) - \lambda \log \left( \frac{|\Sigma_Y|^{1/n}}{|\Sigma_{Y_u}|^{1/n}} \right) \right] \quad (118)$$

$$+ \frac{n}{2} \left[ \log \left( \frac{|\Sigma_{X_u} \Sigma_{Y_u}^{-1} \Sigma_{X_u}|^{1/n} (\lambda - 1)}{|\Sigma_{X_u} - \Sigma_{X_u} \Sigma_{Y_u}^{-1} \Sigma_{X_u}|^{1/n}} \right) - \lambda \log \left( \frac{|\Sigma_{X_u}|^{1/n} (\lambda - 1)}{|\Sigma_{X_u} - \Sigma_{X_u} \Sigma_{Y_u}^{-1} \Sigma_{X_u}|^{1/n} \lambda} \right) \right] = \frac{n}{2} \left[ \log \left( \frac{|\Sigma_X|^{1/n}}{|\Sigma_{X_u}|^{1/n}} \right) - \lambda \log \left( \frac{|\Sigma_Y|^{1/n}}{|\Sigma_{Y_u}|^{1/n}} \right) \right] \quad (119)$$

$$+ \frac{n}{2} \left[ \log \left( \frac{|\Sigma_{X_u}|^{1/n} (\lambda - 1)}{|\Sigma_{Y_u} - \Sigma_{X_u}|^{1/n}} \right) - \lambda \log \left( \frac{|\Sigma_{Y_u}|^{1/n} (\lambda - 1)}{|\Sigma_{Y_u} - \Sigma_{X_u}|^{1/n} \lambda} \right) \right] = \frac{n}{2} \left[ \log \left( \frac{|\Sigma_X|^{1/n} (\lambda - 1)}{|\Sigma_Z|^{1/n}} \right) - \lambda \log \left( \frac{|\Sigma_X + \Sigma_Z|^{1/n} (\lambda - 1)}{|\Sigma_Z|^{1/n} \lambda} \right) \right]. \quad (120)$$

On the other hand, if  $\Sigma_{X_u}$  is such that  $0 \leq \lambda \leq 1 + \frac{|\Sigma_Z|^{1/n}}{|\Sigma_{X_u}|^{1/n}}$ . Then, Lemma 9 allows us to lower bound the integrand in (116) as:

$$h(\mathbf{X}) - h(\mathbf{X}|u) - \lambda(h(\mathbf{Y}) - h(\mathbf{Y}|u)) + I(\mathbf{Y}; V|U = u) - \lambda I(\mathbf{X}; V|U = u) \geq \frac{n}{2} \left[ \log \left( \frac{|\Sigma_X|^{1/n}}{|\Sigma_{X_u}|^{1/n}} \right) - \lambda \log \left( \frac{|\Sigma_Y|^{1/n}}{|\Sigma_{Y_u}|^{1/n}} \right) \right] \quad (121)$$

$$- \frac{\lambda n}{2} \log \left( \frac{|\Sigma_{X_u}|^{1/n}}{|\Sigma_{X_u} \Sigma_{Y_u}^{-1} \Sigma_{X_u}|^{1/n} + |\Sigma_{X_u} - \Sigma_{X_u} \Sigma_{Y_u}^{-1} \Sigma_{X_u}|^{1/n}} \right) = \frac{n}{2} \left[ \log \left( \frac{|\Sigma_X|^{1/n}}{|\Sigma_{X_u}|^{1/n}} \right) - \lambda \log \left( \frac{|\Sigma_Y|^{1/n}}{|\Sigma_{X_u}|^{1/n} + |\Sigma_Z|^{1/n}} \right) \right] \quad (122)$$

$$\geq \frac{n}{2} \left[ \log \left( \frac{|\Sigma_X|^{1/n} (\lambda - 1)}{|\Sigma_Z|^{1/n}} \right) - \lambda \log \left( \frac{|\Sigma_Y|^{1/n} (\lambda - 1)}{|\Sigma_Z|^{1/n} \lambda} \right) \right], \quad (123)$$

where the last line follows since  $-\log(|\Sigma_{X_u}|^{1/n}) + \lambda \log(|\Sigma_{X_u}|^{1/n} + |\Sigma_Z|^{1/n})$  is a monotone decreasing function of  $|\Sigma_{X_u}|^{1/n}$  provided  $\lambda \leq 1 + \frac{|\Sigma_Z|^{1/n}}{|\Sigma_{X_u}|^{1/n}}$ . Thus, setting  $|\Sigma_{X_u}|^{1/n} = \frac{1}{\lambda - 1} |\Sigma_Z|^{1/n}$  only weakens the inequality. ■

The combination of Lemmas 8-10 proves Theorem 1.

#### IV. CLOSING REMARKS

The focus of this paper was on the extremal result asserted by Theorem 1, and not on operational coding problems. However, since the entropy-power-like inequality of Theorem 1 leads to what is arguably the simplest solution for the two-encoder quadratic Gaussian source coding problem (an archetypical problem in network information theory), we have little doubt that it will find other interesting applications. We provided Theorem 2 as one such example. As another example, Theorem 3 can be applied to show that jointly Gaussian auxiliaries exhaust the rate region for multiterminal source coding under logarithmic loss [9] when the sources are Gaussian. This leads to yet another solution for the two-encoder quadratic Gaussian source coding problem, and unifies the two problems under the paradigm of compression under logarithmic loss.

#### ACKNOWLEDGEMENT

The authors acknowledge several conversations with Tsachy Weissman, Kartik Venkat, and Vignesh Ganapathi-Subramanian which led to deeper intuition and insight. The first author also wishes to acknowledge an inspiring discussion with Chandra

Nair that took place at the 2014 International Zurich Seminar on Communications following the presentation of [10]. This conversation and the following exchanges generated the spark which led to a successful proof of Theorem 3, a long-held conjecture of the first author that was first made public on the *Information Theory b-Log* in March, 2013 [11], [12].

## APPENDIX A PROOF OF THEOREM 2

### Converse Part:

Define  $U = f_x(\mathbf{X}_1, \dots, \mathbf{X}_k)$ , and  $V = f_y(\mathbf{Y}_1, \dots, \mathbf{Y}_k)$ , and suppose that  $U, V$  are such that we can determine an ellipsoid  $\mathcal{E}(A_x, b_x)$  having volume bounded by

$$\left( \text{vol}(\mathcal{E}(A_x, b_x)) \right)^{1/n} = \frac{c_n^{1/n}}{|A_x|^{1/n}} \leq c_n^{1/n} \sqrt{n \nu_x |\Sigma_n|^{1/n}} \quad (124)$$

and containing the points  $\{\mathbf{X}_i\}_{i=1}^k$  with probability at least  $1 - \epsilon_n$ , where  $\lim_{n \rightarrow \infty} \epsilon_n = 0$ .

For  $i \in \{1, \dots, k\}$ , define the indicator random variable  $E_i = \mathbf{1}_{\{\mathbf{X}_i \notin \mathcal{E}(A_x, b_x)\}}$ , and note that

$$\begin{aligned} & \left| \mathbb{E} \left[ (A_x \mathbf{X}_i - \mathbb{E}[A_x \mathbf{X}_i | U, V, E_i]) (A_x \mathbf{X}_i - \mathbb{E}[A_x \mathbf{X}_i | U, V, E_i])^T \mid E_i = 0 \right] \right|^{1/n} \\ & \leq \frac{1}{n} \text{Tr} \left( \mathbb{E} \left[ (A_x \mathbf{X}_i - \mathbb{E}[A_x \mathbf{X}_i | U, V, E_i]) (A_x \mathbf{X}_i - \mathbb{E}[A_x \mathbf{X}_i | U, V, E_i])^T \mid E_i = 0 \right] \right) \end{aligned} \quad (125)$$

$$= \frac{1}{n} \mathbb{E} \left[ \|A_x \mathbf{X}_i - \mathbb{E}[A_x \mathbf{X}_i | U, V, E_i]\|^2 \mid E_i = 0 \right] \quad (126)$$

$$\leq \frac{1}{n} \mathbb{E} \left[ \|A_x \mathbf{X}_i - b_x\|^2 \mid E_i = 0 \right] \quad (127)$$

$$\leq \frac{1}{n}, \quad (128)$$

where (125) follows from the inequality of arithmetic and geometric means, (127) follows since conditional expectation minimizes mean square error, and (128) follows since  $\{E_i = 0\} \Rightarrow \{\mathbf{X}_i \in \mathcal{E}(A_x, b_x)\} \Rightarrow \{\|A_x \mathbf{X}_i - b_x\| \leq 1\}$ . Summarizing the above, we have established that, for each  $i = 1, \dots, k$ ,

$$\left| \mathbb{E} \left[ (A_x \mathbf{X}_i - \mathbb{E}[A_x \mathbf{X}_i | U, V, E_i]) (A_x \mathbf{X}_i - \mathbb{E}[A_x \mathbf{X}_i | U, V, E_i])^T \mid E_i = 0 \right] \right|^{1/n} \leq \frac{1}{n}. \quad (129)$$

Applying (129) in conjunction with the maximum-entropy property of Gaussians, we have

$$\frac{1}{n} I(\mathbf{X}_i; U, V | E_i = 0) = \frac{1}{n} h(\mathbf{X}_i | E_i = 0) - \frac{1}{n} h(\mathbf{X}_i | U, V, E_i = 0) \quad (130)$$

$$= \frac{1}{n} h(\mathbf{X}_i | E_i = 0) - \frac{1}{n} h(A_x \mathbf{X}_i | U, V, E_i = 0) + \frac{1}{n} \log |A_x| \quad (131)$$

$$\geq \frac{1}{n} h(\mathbf{X}_i | E_i = 0) - \frac{1}{2n} \log((2\pi e)^n) - \frac{1}{2} \log \left( \frac{1}{n} \right) + \frac{1}{2n} \log \frac{1}{n^n \nu_x^n |\Sigma_n|} \quad (132)$$

$$= \frac{1}{n} \left( h(\mathbf{X}_i | E_i = 0) - h(\mathbf{X}_i) \right) + \frac{1}{2} \log \frac{1}{\nu_x}. \quad (133)$$

Since  $I(\mathbf{X}_i; U, V) \geq I(\mathbf{X}_i; U, V | E_i = 0) - H(E_i) \geq \Pr\{E_i = 0\} I(\mathbf{X}_i; U, V | E_i = 0) - H(E_i)$  and  $\Pr\{E_i = 0\} \geq 1 - \epsilon_n \rightarrow 1$ , it follows by Lemma 13 (see Appendix C) that, for any  $\delta > 0$ ,

$$\frac{1}{n} I(\mathbf{X}_i; U, V) \geq \frac{1}{2} \log \frac{1}{\nu_x \sqrt{1 + \delta}} \quad (134)$$

for  $n$  sufficiently large. Since  $\mathbf{X}_1, \dots, \mathbf{X}_k$  are mutually independent, we have

$$\frac{1}{kn} I(\mathbf{X}_1, \dots, \mathbf{X}_k; U, V) \geq \frac{1}{kn} \sum_{i=1}^k I(\mathbf{X}_i; U, V) \geq \frac{1}{2} \log \frac{1}{\nu_x \sqrt{1 + \delta}} \quad (135)$$

for all  $n$  sufficiently large. By a symmetric argument, it also holds that

$$\frac{1}{kn} I(\mathbf{Y}_1, \dots, \mathbf{Y}_k; U, V) \geq \frac{1}{2} \log \frac{1}{\nu_y \sqrt{1 + \delta}} \quad (136)$$

for  $n$  sufficiently large. Thus, since

$$R_x + R_y \geq \frac{1}{kn} H(U, V) \geq \frac{1}{kn} I(\mathbf{X}_1, \mathbf{Y}_1, \dots, \mathbf{X}_k, \mathbf{Y}_k; U, V), \quad (137)$$

it follows from (4) and (5) that

$$R_x + R_y \geq \frac{1}{2} \log \frac{(1 - \rho^2)\beta((1 + \delta)\nu_x\nu_y)}{2(1 + \delta)\nu_x\nu_y}. \quad (138)$$

Since  $\delta > 0$  can be taken arbitrarily small, (20) must hold.

Likewise, since (2) implies

$$2^{\frac{2}{kn}(I(\mathbf{X}_1, \dots, \mathbf{X}_k; U|V) - I(\mathbf{X}_1, \dots, \mathbf{X}_k; U, V))} = 2^{-\frac{2}{kn}I(\mathbf{X}_1, \dots, \mathbf{X}_k; V)} \geq (1 - \rho^2) + \rho^2 2^{-\frac{2}{kn}I(\mathbf{Y}_1, \dots, \mathbf{Y}_k; V)}, \quad (139)$$

it follows that

$$2^{\frac{2}{kn}(H(U) - I(\mathbf{X}_1, \dots, \mathbf{X}_k; U, V))} \geq (1 - \rho^2) + \rho^2 2^{-\frac{2}{kn}H(V)}. \quad (140)$$

Therefore, (135) also implies (18). By a symmetric argument, we obtain (19).

#### Direct Part:

Fix  $\epsilon > 0$ , and assume (18)-(20) are satisfied. Further, diagonalize  $\Sigma = U_\Sigma \Lambda U_\Sigma^T$  (throughout the proof, we suppress the explicit dependence of the covariance matrices on  $n$  for convenience).

Suppose  $(X^n, Y^n)$  is a sequence of independent pairs of random variables, where  $(X_j, Y_j)$  are jointly normal with linear correlation  $\rho$ , and  $X_i, Y_i$  each have unit variance. For  $n$  sufficiently large, the achievability result for the two-encoder quadratic Gaussian source coding problem (e.g., [2, Theorem 1]) implies that there exist functions  $f_x : \mathbb{R}^n \rightarrow \{1, 2, \dots, 2^{nR_x}\}$  and  $f_y : \mathbb{R}^n \rightarrow \{1, 2, \dots, 2^{nR_y}\}$  for which

$$\Pr \{ \mathbb{E} [\|X^n - \mathbb{E}[X^n|f_x(X^n), f_y(Y^n)]\|^2] > n\nu_x \} < \epsilon \quad \text{and} \quad (141)$$

$$\Pr \{ \mathbb{E} [\|Y^n - \mathbb{E}[Y^n|f_x(X^n), f_y(Y^n)]\|^2] > n\nu_y \} < \epsilon. \quad (142)$$

Therefore, the rates  $(R_x, R_y)$  are sufficient to communicate  $k$  ellipsoids of the form

$$\mathcal{E}_i = \{x : \|A_x x - b_i\| \leq 1\} \quad \text{for } i = 1, 2, \dots, k, \quad (143)$$

where  $A_x \triangleq \frac{1}{\sqrt{n\nu_x}} \Lambda^{-1/2} U_\Sigma^T$ ,  $b_i \triangleq A_x \mathbb{E}[\mathbf{X}_i | f_x(\mathbf{X}_i), f_y(\mathbf{Y}_i)]$ , and  $\mathbf{X}_i \in \mathcal{E}_i$  with probability greater than  $1 - \epsilon$  for  $i = 1, 2, \dots, k$ . This follows since the pair of vectors  $(\Lambda^{-1/2} U_\Sigma^T \mathbf{X}_i, \Lambda^{-1/2} U_\Sigma^T \mathbf{Y}_i)$  is equal in distribution to  $(X^n, Y^n)$ .

Let  $\delta > 0$  satisfy  $\sqrt{\delta/\nu_x} < \epsilon$ , and define  $\tau = 1 - 2\sqrt{\delta}$ , and  $\gamma = (1 - \delta)\tau$  for convenience. Further, let  $u_1, \dots, u_k$  be an orthonormal basis for the vector space spanned by  $b_1, \dots, b_k$ , and define

$$\underline{\mathbf{B}}_{\mathbf{x}} \triangleq \left( \tau I_n - \gamma \sum_{i=1}^k u_i u_i^T \right) A_x. \quad (144)$$

We remark that  $\underline{\mathbf{B}}_{\mathbf{x}}$  is a random matrix, and is a function of  $\{f_x(\mathbf{X}_i), f_y(\mathbf{Y}_i)\}_{i=1}^k$ . Note that

$$\underline{\mathbf{B}}_{\mathbf{x}} \mathbf{X}_i = \left( \tau I_n - \gamma \sum_{i=1}^k u_i u_i^T \right) A_x \mathbf{X}_i = \left( \tau I_n - \gamma \sum_{i=1}^k u_i u_i^T \right) \tilde{\mathbf{X}}_i, \quad (145)$$

where  $\tilde{\mathbf{X}}_i \triangleq A_x \mathbf{X}_i$  for  $i = 1, 2, \dots, k$ . Define  $\mathbf{Z}_i = \tilde{\mathbf{X}}_i - b_i$ , and continue with

$$\underline{\mathbf{B}}_{\mathbf{x}} \mathbf{X}_i = \left( \tau I_n - \gamma \sum_{j=1}^k u_j u_j^T \right) \tilde{\mathbf{X}}_i = \tau(\tilde{\mathbf{X}}_i - b_i) + \tau b_i - \gamma \sum_{j=1}^k u_j u_j^T (\mathbf{Z}_i + b_i) \quad (146)$$

$$= \tau(\tilde{\mathbf{X}}_i - b_i) + (\tau - \gamma)b_i - \gamma \sum_{j=1}^k u_j u_j^T \mathbf{Z}_i, \quad (147)$$

where (147) follows since  $\sum_{j=1}^k u_j u_j^T$  is an orthogonal projection onto the space spanned by  $b_1, \dots, b_k$ . Applying the triangle

inequality, we can conclude

$$\|\underline{\mathbf{B}}_{\mathbf{x}} \mathbf{X}_i\| \leq \tau \|\tilde{\mathbf{X}}_i - b_i\| + (\tau - \gamma) \|b_i\| + \gamma \left\| \sum_{j=1}^k u_j u_j^T \mathbf{Z}_i \right\| \quad (148)$$

$$= \tau \|\tilde{\mathbf{X}}_i - b_i\| + (\tau - \gamma) \|b_i\| + \gamma \sqrt{\sum_{j=1}^k (u_j^T \mathbf{Z}_i)^2} \quad (149)$$

$$\leq \tau \|\tilde{\mathbf{X}}_i - b_i\| + \delta \|b_i\| + \sqrt{\sum_{j=1}^k (u_j^T \mathbf{Z}_i)^2}. \quad (150)$$

Using Jensen and Markov's inequalities, we can bound

$$\Pr \left\{ \|b_i\| \geq \frac{1}{\sqrt{\delta}} \right\} = \Pr \left\{ \left\| \mathbb{E} [\tilde{\mathbf{X}}_i | f_x(\mathbf{X}_i), f_y(\mathbf{Y}_i)] \right\| \geq \frac{1}{\sqrt{\delta}} \right\} \quad (151)$$

$$\leq \Pr \left\{ \mathbb{E} [\|\tilde{\mathbf{X}}_i\| | f_x(\mathbf{X}_i), f_y(\mathbf{Y}_i)] \geq \frac{1}{\sqrt{\delta}} \right\} \quad (152)$$

$$\leq \sqrt{\delta} \mathbb{E} [\|\tilde{\mathbf{X}}_i\|] \quad (153)$$

$$\leq \sqrt{\delta \mathbb{E} [\|\tilde{\mathbf{X}}_i\|^2]} \quad (154)$$

$$= \sqrt{\frac{\delta}{\nu_x}} \quad (155)$$

$$\leq \epsilon, \quad (156)$$

where (155) follows since  $\tilde{X}_i \sim N(0, \frac{1}{n\nu_x} I_n)$ .

Next, since  $b_j$  is the LLMSE estimator of  $\tilde{X}_i$  given  $f_x(\mathbf{X}_i), f_y(\mathbf{Y}_i)$  by construction, we have

$$\mathbb{E} \left[ \sum_{j=1}^k (u_j^T \mathbf{Z}_i)^2 \right] = \sum_{j=1}^k u_j^T \Sigma_{\mathbf{Z}_i} u_j \leq \frac{k}{n\nu_x}, \quad (157)$$

where the inequality follows since the center quantity is upper bounded by the  $k$  largest eigenvalues of  $\Sigma_{\mathbf{Z}_i}$ , which are themselves upper bounded by the  $k$  largest eigenvalues of  $\Sigma_{\tilde{\mathbf{X}}_i} = \frac{1}{n\nu_x} I_n$ . Therefore, proceeding with Markov's inequality, we have

$$\Pr \left\{ \sum_{j=1}^k (u_j^T \mathbf{Z}_i)^2 \geq \delta \right\} \leq \frac{k}{\delta n\nu_x}, \quad (158)$$

which is upper-bounded by  $\epsilon$  for  $n$  sufficiently large.

Also by construction,  $\|\tilde{\mathbf{X}}_i - b_i\| = \|A_x \mathbf{X}_i - b_i\| \leq 1$  with probability greater than  $1 - \epsilon$ . Therefore for  $n$  sufficiently large, we can conclude that

$$\Pr \left\{ \|\underline{\mathbf{B}}_{\mathbf{x}} \mathbf{X}_i\| \leq 1 \right\} > 1 - 3\epsilon. \quad (159)$$

Finally, note that

$$|\underline{\mathbf{B}}_{\mathbf{x}}| = |A_x| \left| \tau I_n - \gamma \sum_{i=1}^k u_i u_i^T \right| = |A_x| \tau^{n-k} (\tau - \gamma)^k, \quad (160)$$

and so

$$|\underline{\mathbf{B}}_{\mathbf{x}}|^{1/n} = \tau \delta^{k/n} |A_x|^{1/n} = \frac{(1 - 2\sqrt{\delta}) \delta^{k/n}}{\sqrt{n\nu_x |\Sigma|^{1/n}}}. \quad (161)$$

Since  $\delta$  can be taken arbitrarily small, a symmetric argument involving the  $\mathbf{Y}_i$ 's completes the proof.

APPENDIX B  
EXISTENCE OF MINIMIZERS

We will require the following result on lower semicontinuity of relative entropy.

**Lemma 11.** [13, Lemma 1.4.3] *Let  $\mathcal{X}$  be a Polish space, and let  $\mathcal{P}(\mathcal{X})$  denote the set of probability measures on  $\mathcal{X}$ . The relative entropy  $D(P\|Q)$  is a lower semicontinuous function of  $(P, Q) \in \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{X})$  with respect to the weak topology.*

**Lemma 12.** *The infimum in (22) is attained.*

*Proof:* We can assume  $\lambda > 1$ , else the data processing inequality implies that  $F(\lambda) \geq 0$ , which is easily attained. First, we show that if  $\{U_n, X_n, Y_n, V_n\}_{n \geq 1}$  is a sequence of candidate minimizers<sup>2</sup> of (22) which converge weakly to  $(U^*, X^*, Y^*, V^*)$ , then  $(U^*, X^*, Y^*, V^*)$  is also a candidate minimizer. To see this, note that Lemma 11 asserts that relative entropy is lower semicontinuous with respect to the weak topology, and hence

$$0 = \lim_{n \rightarrow \infty} D(P_{X_n Y_n} \| P_{X Y}) \geq D(P_{X^* Y^*} \| P_{X Y}) \geq 0. \quad (162)$$

Therefore, we see that  $(X^*, Y^*)$  must be equal to  $(X, Y)$  in distribution. Similarly, by recognizing that (unconditional) mutual information is a relative entropy, lower semicontinuity also yields

$$I(X; Y) = \lim_{n \rightarrow \infty} I(U_n, X_n; Y_n, V_n) \geq I(U^*, X^*; Y^*, V^*) \geq I(X^*; Y^*) = I(X; Y). \quad (163)$$

Hence, we can conclude that  $U^* - X^* - Y^* - V^*$ , and therefore  $(U^*, X^*, Y^*, V^*)$  is a candidate minimizer as desired.

Suppose again that  $\{U_n, X_n, Y_n, V_n\}_{n \geq 1}$  is a sequence of candidate minimizers that converges weakly to  $(U^*, X^*, Y^*, V^*)$ . As established previously,  $(X_n, Y_n) = (X^*, Y^*) = (X, Y)$  in distribution. Fix an arbitrary conditional distribution  $P_{Z|Y}$ . Let  $P_Y \rightarrow P_{Z|Y} \rightarrow P_Z$ . By lower semicontinuity of relative entropy, we have:

$$\liminf_{n \rightarrow \infty} D(P_{X_n Y_n U_n} P_Z \| P_{X_n U_n} P_{Y Z}) \geq D(P_{X^* Y^* U^*} P_Z \| P_{X^* U^*} P_{Y Z}) \quad (164)$$

Thus, there exists  $\delta(n) \rightarrow 0$  such that

$$D(P_{X_n Y_n U_n} P_Z \| P_{X_n U_n} P_{Y Z}) \geq D(P_{X^* Y^* U^*} P_Z \| P_{X^* U^*} P_{Y Z}) - \delta(n), \quad (165)$$

which implies

$$\inf_{P_{Z|Y}} D(P_{X_n Y_n U_n} P_Z \| P_{X_n U_n} P_{Y Z}) \geq \inf_{P_{Z|Y}} D(P_{X^* Y^* U^*} P_Z \| P_{X^* U^*} P_{Y Z}) - \delta(n). \quad (166)$$

By the variational representation of mutual information, the infima on the LHS and RHS are attained when  $P_{Z|Y} = P_{U_n|Y}$  and  $P_{Z|Y} = P_{U^*|Y}$ , respectively. Therefore,

$$I(X_n; Y_n | U_n) \geq I(X^*; Y^* | U^*) - \delta(n). \quad (167)$$

Since  $\delta(n)$  vanishes, we can conclude that

$$\liminf_{n \rightarrow \infty} I(X_n; Y_n | U_n) \geq I(X^*; Y^* | U^*). \quad (168)$$

Now, we can add  $2\lambda I(X; Y)$  to the functional being considered in (22) without changing the nature of the optimization problem. Therefore, we aim to show that the infimum of the functional

$$I(X; U) - \lambda I(Y; U) + I(Y; V | U) - \lambda I(X; V | U) + 2\lambda I(X; Y) \quad (169)$$

$$= I(X; U) + \lambda I(X; Y | U) + I(Y; V) + \lambda I(X; Y | V) + (\lambda - 1)I(U; V). \quad (170)$$

is attained. By our previous observation, if  $\{U_n, X_n, Y_n, V_n\}_{n \geq 1}$  is a sequence of candidate minimizers which approach the infimum of (22) and converge weakly to  $(U^*, X^*, Y^*, V^*)$ , then we can apply lower semicontinuity again together with (168) and its symmetric variant  $\liminf_{n \rightarrow \infty} I(X_n; Y_n | V_n) \geq I(X^*; Y^* | V^*)$  to obtain

$$\begin{aligned} F(\lambda) + 2\lambda I(X; Y) &= \lim_{n \rightarrow \infty} I(X_n; U_n) + \lambda I(X_n; Y_n | U_n) + I(Y_n; V_n) + \lambda I(X_n; Y_n | V_n) + (\lambda - 1)I(U_n; V_n) \\ &\geq I(X^*; U^*) + \lambda I(X^*; Y^* | U^*) + I(Y^*; V^*) + \lambda I(X^*; Y^* | V^*) + (\lambda - 1)I(U^*; V^*) \end{aligned} \quad (171)$$

$$\geq F(\lambda) + 2\lambda I(X; Y), \quad (172)$$

implying equality throughout, and optimality of  $(U^*, X^*, Y^*, V^*)$ .

<sup>2</sup>i.e.,  $(X_n, Y_n)$  equal  $(X, Y)$  in distribution, and  $U_n - X_n - Y_n - V_n$  for each  $n$ .



Thus, we only need to show that if  $\{U_n, X_n, Y_n, V_n\}_{n \geq 1}$  is a sequence of candidate minimizers, then there exists a subsequence  $\{U_{n_k}, X_{n_k}, Y_{n_k}, V_{n_k}\}_{k \geq 1}$  which converges weakly to some  $(U^*, X^*, Y^*, V^*)$ . Since mutual information is invariant under one-to-one transformations of support, we can assume without loss of generality that  $U_n, V_n$  are each supported on the interval  $[0, 1]$  for all  $n$ . Recalling Prohorov's theorem [14, Theorem 3.9.2], we only need to show that the sequence of measures  $P_n$  is tight, where  $P_n$  is the joint distribution of  $(U_n, X_n, Y_n, V_n)$ . To this end, note that for any  $\epsilon > 0$ , we can choose  $t$  sufficiently large so that

$$P_n \{(U_n, X_n, Y_n, V_n) \notin [-t, t]^4\} = P_n \{(X_n, Y_n) \notin [-t, t]^2\} < \epsilon. \quad (173)$$

Thus, the claim is proved.  $\blacksquare$

## APPENDIX C AUXILIARY LEMMAS

**Lemma 13.** For  $n = 1, 2, \dots$ , suppose  $X^n \sim N(0, \Sigma_n)$ , where  $\Sigma_n \in \mathbb{R}^{n \times n}$  is positive definite, and let  $E_n \in \{0, 1\}$  be correlated with  $X^n$ . If  $\lim_{n \rightarrow \infty} \Pr\{E_n = 0\} = 1$ , then

$$\lim_{n \rightarrow \infty} \frac{1}{n} \left( h(X^n | E_n = 0) - h(X^n) \right) = 0. \quad (174)$$

*Proof:* For the proof, we suppress the explicit dependence of  $X^n$  and  $\Sigma_n$  on  $n$ , and simply write  $\mathbf{X}$  and  $\Sigma$ . Note that

$$H(E_n) \geq I(\mathbf{X}; E_n) = \Pr\{E_n = 0\} D(P_{\mathbf{X}|E_n=0} \| P_{\mathbf{X}}) + \Pr\{E_n = 1\} D(P_{\mathbf{X}|E_n=1} \| P_{\mathbf{X}}) \quad (175)$$

$$\geq \Pr\{E_n = 0\} D(P_{\mathbf{X}|E_n=0} \| P_{\mathbf{X}}), \quad (176)$$

and therefore  $D(P_{\mathbf{X}|E_n=0} \| P_{\mathbf{X}}) \rightarrow 0$ . Define  $Z_n = \frac{1}{n} \mathbf{X}^T \Sigma^{-1} \mathbf{X}$ , and  $W_n = \frac{1}{n} \mathbf{X}^T \Sigma^{-1} \mathbf{X} | \{E_n = 0\}$ . By the data processing theorem for relative entropy,  $D(P_{W_n} \| P_{Z_n}) \rightarrow 0$  as well. Thus, for any  $\epsilon > 0$ , Pinsker's inequality and the WLLN together imply

$$\Pr\{|W_n - 1| \geq \epsilon\} = \Pr\{|W_n - 1| \geq \epsilon\} - \Pr\{|Z_n - 1| \geq \epsilon\} + \Pr\{|Z_n - 1| \geq \epsilon\} \quad (177)$$

$$\leq \|P_{W_n} - P_{Z_n}\|_{TV} + \Pr\{|Z_n - 1| \geq \epsilon\} \quad (178)$$

$$\leq \epsilon \quad (179)$$

for all  $n$  sufficiently large, establishing that  $W_n \rightarrow 1$  in probability. For any function  $f: \mathbb{R}^n \rightarrow \mathbb{R}$

$$\mathbb{E} [\mathbf{X}^T \Sigma^{-1} \mathbf{X} f(\mathbf{X})] = \Pr\{E_n = 0\} \mathbb{E} [\mathbf{X}^T \Sigma^{-1} \mathbf{X} f(\mathbf{X}) | E_n = 0] + \Pr\{E_n = 1\} \mathbb{E} [\mathbf{X}^T \Sigma^{-1} \mathbf{X} f(\mathbf{X}) | E_n = 1],$$

so it follows that

$$\Pr\{E_n = 0\} \mathbb{E} [W_n 1_{\{W_n \geq K\}}] \leq \mathbb{E} [Z_n 1_{\{Z_n \geq K\}}] \quad (180)$$

by non-negativity of  $W_n$  and  $Z_n$ . The Cauchy-Schwarz and Markov inequalities together imply

$$|\mathbb{E} [Z_n 1_{\{Z_n \geq K\}}]|^2 \leq \mathbb{E} [Z_n^2] \Pr\{Z_n \geq K\} = \frac{n+2}{n} \Pr\{Z_n \geq K\} \leq \frac{3}{K}, \quad (181)$$

and therefore (180) implies that  $\{W_n\}_{n \geq 1}$  is uniformly integrable. It follows that  $\mathbb{E}[W_n] \rightarrow 1 = \mathbb{E}[Z_n]$  (e.g., [14, Theorem 5.5.2]).

To conclude, we observe that

$$\frac{1}{n} \left( h(\mathbf{X} | E_n = 0) - h(\mathbf{X}) \right) = -\frac{1}{n} D(P_{\mathbf{X}|E_n=0} \| P_{\mathbf{X}}) + \frac{1}{n} \int_{\mathbb{R}^n} \left( P_{\mathbf{X}}(\mathbf{x}) - P_{\mathbf{X}|E_n=0}(\mathbf{x}) \right) \log(P_{\mathbf{X}}(\mathbf{x})) d\mathbf{x} \quad (182)$$

$$= -\frac{1}{n} D(P_{\mathbf{X}|E_n=0} \| P_{\mathbf{X}}) - \frac{1}{2n} \mathbb{E} [\mathbf{X}^T \Sigma^{-1} \mathbf{X}] + \frac{1}{2n} \mathbb{E} [\mathbf{X}^T \Sigma^{-1} \mathbf{X} | E_n = 0] \quad (183)$$

$$= -\frac{1}{n} D(P_{\mathbf{X}|E_n=0} \| P_{\mathbf{X}}) + \frac{1}{2} (\mathbb{E}[W_n] - \mathbb{E}[Z_n]), \quad (184)$$

which completes the proof.  $\blacksquare$

Lemma 13 can be viewed as a regularity property enjoyed by Gaussian vectors. Though not needed elsewhere in this paper, it is interesting to note that Lemma 13 is sharp in the following sense:

**Proposition 1.** For  $n = 1, 2, \dots$ , let  $X^n \sim N(0, \Sigma_n)$ , where  $\Sigma_n \in \mathbb{R}^{n \times n}$  is positive definite. For any function  $g(n) \rightarrow \infty$  arbitrarily slowly, there exists a sequence of random variables  $E_n \in \{0, 1\}$  such that  $\lim_{n \rightarrow \infty} \Pr\{E_n = 0\} = 1$  and

$$\lim_{n \rightarrow \infty} \frac{g(n)}{n} \left( h(X^n | E_n = 0) - h(X^n) \right) = \infty. \quad (185)$$

In particular, the normalization by  $1/n$  in (174) is essential.

*Proof:* From the proof of Lemma 13, we can assume without loss of generality that  $X^n$  is iid  $N(0, 1)$ . Let  $E_n = 0$  if there are at least  $f(n)$  different  $X_i$ 's for which  $|X_i| \geq \sqrt{2}$ , and let  $E_n = 1$  otherwise.

Since the  $X_i$ 's are independent, we can see that

$$\mathbb{E} \left[ \sum_i X_i^2 \middle| E_n = 0 \right] \geq 2f(n) + (n - f(n)) = f(n) + n. \quad (186)$$

On the other hand,

$$\mathbb{E} \left[ \sum_i X_i^2 \right] = n. \quad (187)$$

Now, we have that  $\Pr\{E_n = 0\} = \Pr\{B(n, p) \geq f(n)\}$ , where  $B(n, p)$  is a Binomial random variable consisting of  $n$  trials with bias  $p = \Pr\{|X_i| \geq \sqrt{2}\}$ . Continuing, we have

$$\Pr\{E_n = 0\} = \Pr\{B(n, p) \geq f(n)\} = \Pr \left\{ \frac{1}{\sqrt{n}}(B(n, p) - np) \geq \sqrt{n} \left( \frac{1}{n}f(n) - p \right) \right\}. \quad (188)$$

By the CLT,  $\frac{1}{\sqrt{n}}(B(n, p) - np) \rightarrow N(0, p(1-p))$  in distribution. Therefore,  $\Pr\{E_n = 0\} \rightarrow 1$  provided  $f(n) = o(n)$ . Recalling the proof of Lemma 13, we have

$$h(X^n | E_n = 0) - h(X^n) \geq \frac{1}{2}f(n) + o(1). \quad (189)$$

Thus, the claim is proved by putting  $f(n) = n/\sqrt{g(n)}$ , where  $g(n) \rightarrow \infty$  arbitrarily slowly. ■

**Lemma 14.** Consider a function  $f : [0, \infty) \rightarrow \mathbb{R}$  defined implicitly by:

$$2^{-2t} = a_1 2^{-2f(t)} + a_2. \quad (190)$$

If  $a_1 + a_2 \leq 1$ , then

$$\min_{t \geq 0} \left\{ \max\{f(t), 0\} - \lambda t \right\} = \begin{cases} \frac{1}{2} \log \left( \frac{a_1(\lambda-1)}{a_2} \right) - \frac{\lambda}{2} \log \left( \frac{\lambda-1}{a_2 \lambda} \right) & \text{if } \lambda \geq \frac{a_1+a_2}{a_1} \\ -\frac{\lambda}{2} \log \left( \frac{1}{a_1+a_2} \right) & \text{if } 0 \leq \lambda \leq \frac{a_1+a_2}{a_1}. \end{cases} \quad (191)$$

*Proof:* Note that  $f'(t) = \frac{2^{-2t}}{2^{-2t}-a_2}$ , and therefore  $f'(t) = \lambda \Rightarrow t = \frac{1}{2} \log \frac{\lambda-1}{a_2 \lambda}$ . Now, suppose  $a_1 + a_2 \leq 1$ . Then  $f(t) = 0 \Rightarrow t = \frac{1}{2} \log \frac{1}{a_1+a_2} \geq 0$ . Define  $f_+(t) = \max\{f(t), 0\}$ . Like  $f(t)$ ,  $f_+(t)$  is a convex increasing function. For  $\lambda > f' \left( \frac{1}{2} \log \frac{1}{a_1+a_2} \right) = \frac{a_1+a_2}{a_1}$ ,  $\lambda$  is a derivative of  $f_+(t)$  at  $t = \frac{1}{2} \log \frac{\lambda-1}{a_2 \lambda}$ . On the other hand, if  $0 \leq \lambda \leq \frac{a_1+a_2}{a_1}$ , then  $\lambda$  is a subderivative of  $f_+(t)$  at  $t = \frac{1}{2} \log \frac{1}{a_1+a_2}$ . Therefore, we can conclude that

$$\min_{t \geq 0} \{f_+(t) - \lambda t\} = \begin{cases} \frac{1}{2} \log \left( \frac{a_1(\lambda-1)}{a_2} \right) - \frac{\lambda}{2} \log \left( \frac{\lambda-1}{a_2 \lambda} \right) & \text{if } \lambda \geq \frac{a_1+a_2}{a_1} \\ -\frac{\lambda}{2} \log \left( \frac{1}{a_1+a_2} \right) & \text{if } 0 \leq \lambda \leq \frac{a_1+a_2}{a_1}. \end{cases} \quad (192)$$

## REFERENCES

- [1] Y. Oohama, "Gaussian multiterminal source coding," *Information Theory, IEEE Transactions on*, vol. 43, pp. 1912–1923, Nov 1997.
- [2] A. Wagner, S. Tavildar, and P. Viswanath, "Rate region of the quadratic gaussian two-encoder source-coding problem," *Information Theory, IEEE Transactions on*, vol. 54, pp. 1938–1961, May 2008.
- [3] M. Henk, "Löwner–John ellipsoids," *Optimization Stories, Documenta Mathematica*, pp. 95–106, 2012.
- [4] S. P. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.
- [5] Y. Geng and C. Nair, "The capacity region of the two-receiver gaussian vector broadcast channel with private and common messages," *Information Theory, IEEE Transactions on*, vol. 60, pp. 2087–2104, April 2014.
- [6] S. M. Srivastava, *A course on Borel sets*, vol. 180. Springer, 1998.
- [7] S. Ghurye, I. Olkin, *et al.*, "A characterization of the multivariate normal distribution," *The Annals of Mathematical Statistics*, vol. 33, no. 2, pp. 533–541, 1962.
- [8] M. Marcus and H. Minc, *A survey of matrix theory and matrix inequalities*, vol. 14. Courier Dover Publications, 1992. p. 115.
- [9] T. Courtade and T. Weissman, "Multiterminal source coding under logarithmic loss," *Information Theory, IEEE Transactions on*, vol. 60, pp. 740–761, Jan 2014.
- [10] T. Courtade, J. Jiao, and T. Weissman, "On an extremal data processing inequality for long Markov chains," in *International Zurich Seminar on Communications (IZS)*, pp. 1 – 4, Feb. 2014.
- [11] T. Courtade, "An extremal conjecture: Experimenting with online collaboration," *Information Theory b-Log*, March 5, 2013. <http://blogs.princeton.edu/blogit/2013/03/05/>.
- [12] T. Javidi, Ed., "In the blogosphere...", *IEEE Information Theory Society Newsletter*, vol. 63, pp. 2087–2104, June 2013.
- [13] P. Dupuis and R. S. Ellis, *A weak convergence approach to the theory of large deviations*, vol. 902. John Wiley & Sons, 2011.
- [14] R. Durrett, *Probability: theory and examples*, vol. 3. Cambridge university press, 2010.